RESEARCH ARTICLE SUMMARY

CRISPR

Programmable gene insertion in human cells with a laboratory-evolved CRISPR-associated transposase

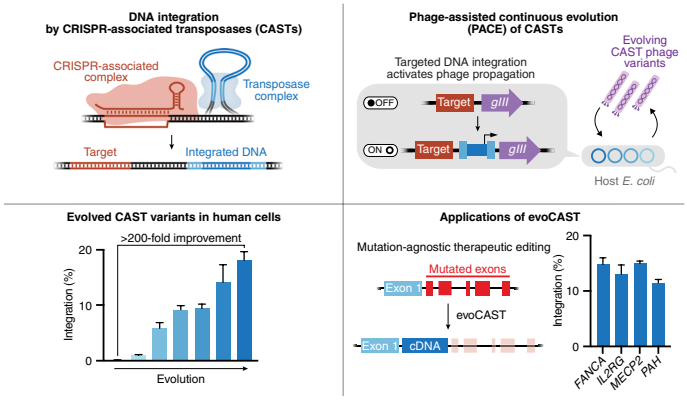
Isaac P. Witte†, George D. Lampe†, Simon Eitzinger†, Shannon M. Miller, Kiara N. Berríos, Amber N. McElroy, Rebeca T. King, Olivia G. Stringham, Diego R. Gelsinger, Phuc Leo H. Vo, Albert T. Chen, Jakub Tolar, Mark J. Osborn, Samuel H. Sternberg*, David R. Liu*

Full article and list of author affiliations: https://doi.org/10.1126/science.adt5199

INTRODUCTION: The efficient insertion of gene-sized DNA sequences at user-specified genomic sites is a long-standing goal in genome editing. Although current editing methods can correct most disease-causing mutations, the genetic diversity underlying many disorders will require the design and regulatory approval of many mutation-specific strategies—substantially limiting the number of patients who can benefit from therapeutic genome editing. Programmed genomic integration of a healthy gene copy could offer a mutation-agnostic treatment for loss-of-function genetic diseases. Additionally, targeted gene integration enables other applications, including cancer immunotherapies, transgenic cell and animal models for basic research, and metabolic engineering.

RATIONALE: CRISPR-associated transposases (CASTs) are naturally occurring bacterial systems that exploit nuclease-deficient CRISPR machinery to integrate DNA at genomic locations specified by guide RNAs. CASTs offer many attractive qualities as a genome editing tool, including facile programmability, compatibility with multi-kilobase-scale DNA cargo, and avoidance of genomic double-strand DNA breaks. Despite this promise, wild-type CASTs reported to date support minimal integration in human cells (often ≤0.1% of treated cells). We reasoned that this low efficiency may stem from naturally evolved, suboptimal transposition catalysis that mitigates mobilization-induced fitness cost to the host. To enable efficient CAST integration in human cells, we developed a phage-assisted continuous evolution (PACE) system that rapidly evolves CAST variants capable of fast targeted transposition and applied CAST-PACE to a prototypical Type I-F CAST system from *Pseudoalteromonas*.

RESULTS: We linked on-target DNA integration in *Escherichia coli* to the propagation of continuously mutating phage genomes encoding evolving CAST components. After hundreds of generations of continuous selection, replication, and mutation in which the resulting phage survived an overall 10^{322}-fold dilution, we generated an evolved variant of the CAST transposase protein TnsB that mediated >200-fold improved integration activity in human cells. The evolved TnsB contains 10 activity-enhancing mutations located throughout the protein, which likely modulate several distinct interactions with other CAST components. Notably, the evolved TnsB mediated efficient integration activity in human cells without requiring codelivery of the bacterial CAST accessory protein, ClpX, which is cytotoxic. We combined this evolved TnsB with other PACE-evolved and rationally engineered CAST components to yield evoCAST, a system optimized for human-cell integration activity. EvoCAST achieved 10 to 30% integration efficiencies across 14 genomic targets in human cells, representing a 420-fold average improvement over wild-type CAST. EvoCAST supported large DNA cargoes >10 kb and mediated the integration of several therapeutic payloads at disease-relevant genomic sites, including safe harbor loci, sites for cancer immunotherapy engineering, and genes implicated in loss-of-function genetic diseases. EvoCAST also performed targeted integration in multiple human cell types, including primary human fibroblasts, and exhibited high product purity, with no detected insertions and deletions (indels), predominantly unidirectional cargo insertion, single–base pair precision of integration, and low levels of off-target integration.



PACE-evolved CASTs mediate efficient, programmable gene integration in human cells. CASTs exploit a nuclease-deficient CRISPR-Cas system to recruit a transposase complex that catalyzes DNA integration (top left). We developed a PACE selection that links integration to propagation of phage that encode transposase components (top right), yielding an evolved CAST that supports >200-fold improved integration in human cells (bottom left). EvoCAST enables efficient and targeted integration of therapeutically relevant genes at many genomic loci (bottom right), paving the way for mutation-agnostic therapies for loss-of-function genetic diseases.

CONCLUSION: This work establishes CAST as a powerful platform technology for efficient, RNA-guided gene integration in human cells. The advantages of evoCAST—including its simple programmability, single-step integration mechanism, and avoidance of genomic double-strand breaks—make it well-suited for many applications in the life sciences and therapeutics, including the capability to address genetically diverse patient populations through a single editing agent. The CAST PACE system developed in this work also provides a strategy for improving the properties of other naturally occurring CASTs toward their use for efficient human-cell genome editing. □

*Corresponding author. Email: drliu@fas.harvard.edu (D.R.L.), shsternberg@gmail.com (S.H.S.) †These authors contributed equally to this work. **Cite this article as** I. P. Witte *et al.*, *Science* **388**, eadt5199 (2025). DOI: 10.1126/science.adt5199

CRISPR

Programmable gene insertion in human cells with a laboratory-evolved CRISPR-associated transposase

Isaac P. Witte[1,2,3]†, George D. Lampe[4,5]†, Simon Eitzinger[1,2,3]†, Shannon M. Miller[1,2,3]‡, Kiara N. Berríos[1,2,3], Amber N. McElroy[6], Rebeca T. King[4], Olivia G. Stringham[1,2,3], Diego R. Gelsinger[4], Phuc Leo H. Vo[4]§, Albert T. Chen[1,2,3], Jakub Tolar[6], Mark J. Osborn[6], Samuel H. Sternberg[4,5]*, David R. Liu[1,2,3]*

Programmable gene integration in human cells has the potential to enable mutation-agnostic treatments for loss-of-function genetic diseases and facilitate many applications in the life sciences. CRISPR-associated transposases (CASTs) catalyze RNA-guided DNA integration but thus far demonstrate minimal activity in human cells. Using phage-assisted continuous evolution (PACE), we generated CAST variants with >200-fold average improved integration activity. The evolved CAST system (evoCAST) achieves ~10 to 30% integration efficiencies of kilobase-size DNA cargoes in human cells across 14 tested genomic target sites, including safe harbor loci, sites used for immunotherapy, and genes implicated in loss-of-function diseases, with undetected indels and low levels of off-target integration. Collectively, our findings establish a platform for the laboratory evolution of CASTs and advance a versatile system for programmable gene integration in living systems.

Advances in programmable nucleases (*1*–*4*), base editors (*5*–*8*), and prime editors (*9*–*12*) have enabled the disruption, installation, or correction of virtually any specified genomic DNA sequence <200 base pairs (bp) in size. These technologies have been effectively deployed in the clinic as one-time treatments for various genetic disorders (*13*–*16*), with more than 60 clinical trials underway (*17*).

Despite this progress, the targeted insertion of gene-sized (≥1 kb) DNA sequences into specified genomic sites in mammalian cells remains a long-standing challenge in genome editing and gene therapy. The mutational heterogeneity underlying many genetic diseases (*18*), such as cystic fibrosis, Stargardt disease, and hemophilia B, complicates maximizing the fraction of patients that can benefit from therapeutic genome editing. Individual nuclease, base editing, and prime editing approaches that target pathogenic alleles typically cannot benefit patients with other mutations in the same gene, necessitating the development and regulatory approval of many different genome editing strategies to treat diverse patient cohorts.

Traditional gene addition therapies use viruses to provide healthy gene copies that rescue loss-of-function mutations (*19*), enabling a single-treatment strategy for many mutations in the same gene. Although effectively used in clinical applications (*19*), viral gene therapies face limitations, including risks of oncogenic DNA integration (which can be attenuated by vector design) (*19*, *20*), potential need for redosing (*21*), and immune responses to viral vectors (*22*). Moreover, genes expressed exogenously or from ectopic genomic loci lack their native regulatory contexts, which can lead to underdosing, overdosing, silencing, or dysregulated function (*23*–*25*).

Programmable insertion of large DNA sequences at endogenous genomic sites could enable one-time, permanent, mutation-agnostic therapies for loss-of-function diseases through the installation of a healthy gene copy at the native locus or a safe-harbor locus. Additionally, programmable DNA insertion could facilitate many other therapeutic and life sciences applications, including the streamlined production of cancer immunotherapies requiring transgenes [e.g., chimeric antigen receptor (CAR) T cell therapy] and the simplified generation of transgenic cell lines and model organisms requiring large payloads.

Nucleases such as CRISPR-Cas9 generate targeted DNA double-strand breaks (DSBs) that can stimulate incorporation of exogenous donor DNA through homology-directed repair (HDR) (*2*, *3*, *26*) or end-joining pathways [e.g., homology-independent targeted integration (HITI)] (*27*). However, HDR requires cellular machinery typically only expressed in dividing cells (*28*), preventing its efficient application in most therapeutically relevant cell types. Although HITI can occur in nondividing cells, integration events lack both orientation and copy number control (*27*). Additionally, DSBs lead to uncontrolled formation of insertions and deletions (indels) at rates comparable to or higher than that of desired DNA integration (*4*, *27*) and are associated with undesired cellular consequences, including chromosomal translocations (*29*), large deletions (*29*), and p53 activation (*30*).

Engineered fusions of transposase and recombinase domains to Cas9 can support DNA integration without requiring DSB formation, but thus far they have shown low efficiency at genomic loci in human cells and frequent off-target integration (*31*–*34*). The combination of prime editing and site-specific recombinases (PASSIGE) (*35*–*37*) can mediate the efficient targeted installation of recombinase attachment sites followed by recombinase-mediated cargo gene insertion. This approach, however, requires coordinated prime editing and recombinase systems to catalyze multiple successive enzymatic steps, which can generate undesired by-products such as indels and attachment sites lacking cargo gene insertion (*35*–*37*). Developing programmable DNA insertion strategies that avoid genomic DSB formation, offer high product purity, and proceed in a single enzymatic step would complement existing approaches and potentially enable new research and therapeutic applications.

CRISPR-associated transposases (CASTs) are recently discovered bacterial systems that use RNA-guided, nuclease-deficient CRISPR-Cas systems to direct kilobase-scale transposon insertion by Tn7-like transposases (*38*–*40*) (Fig. 1A). Tn7-like transposons have exapted multiple distinct CRISPR-Cas subtypes, with type I-F and type V-K CASTs comprising the most extensively characterized systems to date. Type I-F CASTs are especially promising for genome editing applications, exhibiting high insertion efficiency, high on-target specificity, high directionality bias, high product purity, and low incidence of tandem-insertion by-products in *Escherichia coli* (*39*, *41*–*43*).

Despite robust efficiency in bacteria, type I-F CASTs reported to date are minimally active in human cells (*44*). Assessment of diverse type I-F CASTs in human embryonic kidney (HEK) 293T cells identified a *Pseudoalteromonas* sp. S983 system (*Pse*CAST, from Tn*7016*) with <~0.1% genomic DNA insertion efficiency, which improved to ~1% efficiency when supplemented with the bacterial unfoldase ClpX, albeit with increased cytotoxicity (*44*). Although the low activity of *Pse*CAST in human cells could arise from many potential explanations, we reasoned that insertion efficiency might be limited by transposition catalysis or DNA binding, which may have naturally evolved suboptimally to mitigate host fitness costs from excessive transposition (*45*).

[1]Merkin Institute of Transformative Technologies in Healthcare, Broad Institute of Harvard and MIT, Cambridge, MA, USA. [2]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, USA. [3]Howard Hughes Medical Institute, Harvard University, Cambridge, MA, USA. [4]Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY, USA. [5]Howard Hughes Medical Institute, Columbia University, New York, NY, USA. [6]Department of Pediatrics, University of Minnesota Medical School, Minneapolis, MN, USA. ***Corresponding author. Email: drliu@fas.harvard.edu (D.R.L.); shsternberg@gmail.com (S.H.S.)** †These authors contributed equally to this work. ‡Present address: Department of Chemistry, The Scripps Research Institute, La Jolla, CA, USA. §Present address: Vertex Pharmaceuticals, Boston, MA, USA.

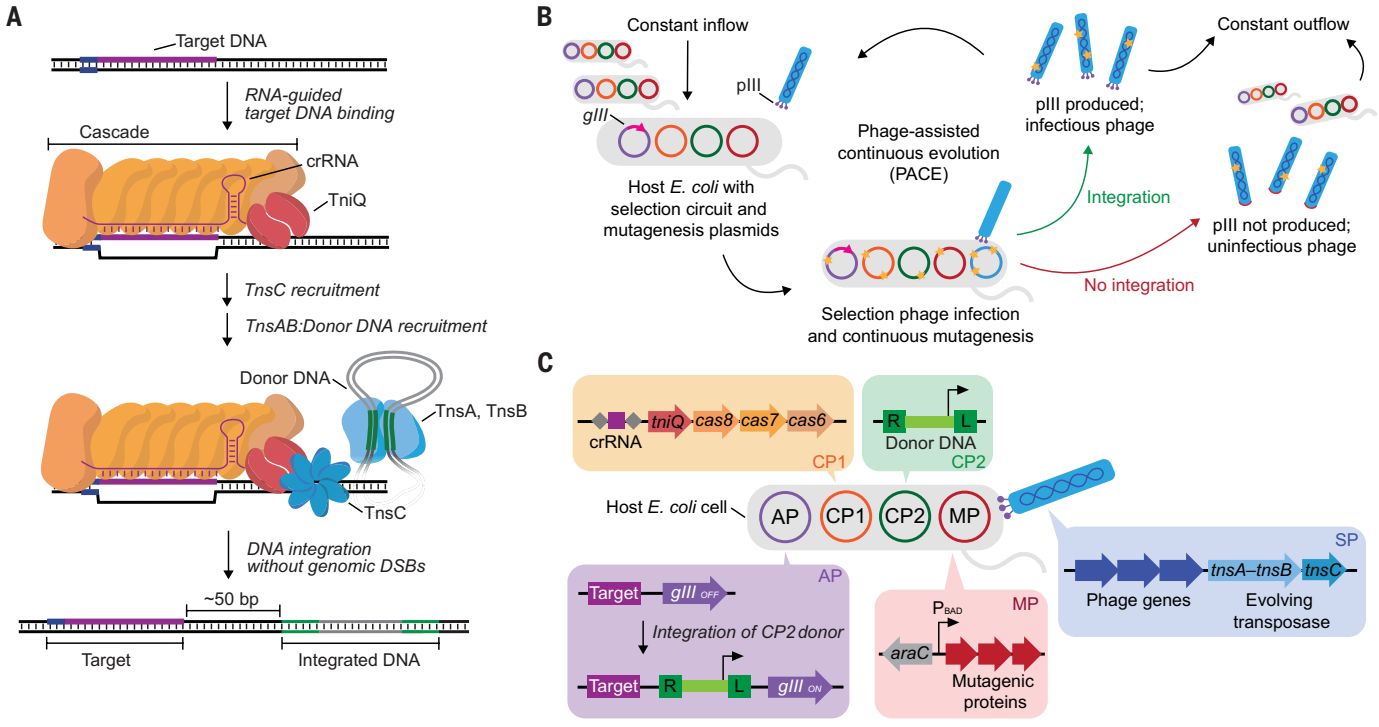


Fig. 1. Phage-assisted continuous evolution (PACE) of CRISPR-associated transposases (CASTs). (**A**) Overview of RNA-guided DNA integration by type I-F CAST. DNA targeting is mediated by the CRISPR effector complex Cascade, comprising Cas6, Cas7, Cas8, and a CRISPR RNA (crRNA) complexed with the transposition protein TniQ (together referred to as QCascade). Target DNA-bound QCascade recruits the AAA+ ATPase TnsC, which subsequently recruits the heteromeric TnsA-TnsB transposase to catalyze excision of the transposon DNA and integration of the transposon at the target locus. (**B**) Overview of PACE for CAST evolution. Selection phage (SP) encodes evolving CAST proteins. Host *E. coli* encode a selection circuit that links CAST integration to *gIII* expression, which produces the essential phage protein pIII. Production of pIII enables SPs encoding active CAST proteins to replicate. PACE occurs in a fixed-volume vessel (the "lagoon") under constant dilution with fresh host *E. coli*, such that only SPs propagating faster than the rate of dilution can persist and evolve. (**C**) Anatomy of the initial CAST PACE selection circuit. An SP encodes evolving transposase proteins TnsA-TnsB [an artificial fusion generated in (*44*)] and TnsC, whereas nonevolving CAST components are encoded on a complementary plasmid (CP1). Integration of a transposon provided on a second complementary plasmid (CP2) into a crRNA-specified target site on the accessory plasmid (AP) installs a promoter upstream of *gIII*, resulting in *gIII* expression and SP propagation. Replicating SPs accumulate mutations induced by a mutagenesis plasmid (MP) (*48*), such that progeny SPs encode new CAST protein variants for selection in subsequent generations.

In this study, we report the application of phage-assisted continuous evolution (PACE) (*46*) to evolve CAST systems that function efficiently in human cells. We evolved the *Pse*CAST transposase module toward increased catalysis through hundreds of generations of mutation, selection, and replication, yielding an evolved transposase variant with >200-fold average improved integration activity in human cells compared with that of wild-type *Pse*CAST, without requiring ClpX. Structure-guided engineering of the DNA-targeting module of *Pse*CAST further improved efficiency, synergizing with the evolved transposase to yield an optimized, evolved CAST (evoCAST) system. EvoCAST supported 10 to 30% insertion efficiencies of kilobase-sized DNA cargoes at 14 tested genomic loci in HEK293T cells, with similar efficiencies in primary human fibroblasts. EvoCAST retained favorable aspects of wild-type *Pse*CAST integration, including high regiospecificity, near unidirectional insertion, and undetected genomic indel formation. Collectively, these results establish a platform for the evolution of CAST systems toward increased activity in mammalian cells and represent a milestone in the development of CASTs for targeted, DSB-free DNA insertion with therapeutically relevant efficiencies.

Results

Development of CAST PACE

PACE maps the key steps of traditional, stepwise directed evolution onto the M13 bacteriophage life cycle, accelerating the laboratory evolution of biomolecules by >100-fold with minimal researcher intervention (*46*, *47*) (Fig. 1B). During PACE, a selection phage (SP) expresses an evolving gene of interest in place of *gIII*, an essential gene for phage replication. *gIII* is instead encoded on an accessory plasmid (AP) in host *E. coli* under a transcriptional circuit linking *gIII* expression to the desired activity of the protein of interest. SP populations are mutagenized through an inducible mutagenesis plasmid (MP) (*48*) and diluted with fresh cells, either continuously (PACE) or periodically [phage-assisted noncontinuous evolution (PANCE)] (*47*), in fixed-volume "lagoons." PACE has been used to evolve many proteins of diverse function, including polymerases (*46*, *49*–*51*), proteases (*52*–*54*), protein-binding proteins (*55*–*57*), DNA-binding proteins (*58*), degrons (*59*), metabolic enzymes (*60*), and genome editing agents (*10*, *61*–*70*).

PACE efficiently and unbiasedly explores vast sequence spaces, and it exhibits few requirements beyond the evolving protein's ability to induce *gIII* expression in *E. coli*. These aspects make PACE well-suited for evolving type I-F CASTs, which are large, multicomponent systems lacking extensive structural and biochemical characterization, limiting rational engineering. Motivated by our hypothesis that integration catalysis may bottleneck activity in human cells, we focused evolution on the transposase module of *Pse*CAST (*43*, *44*) by encoding TnsA, TnsB, and TnsC (referred to hereafter as TnsABC) on the SP.

To evolve TnsABC for increased integration efficiency, we developed a PACE selection linking transposition activity to phage propagation (Fig. 1C). Host cells contain a complementary plasmid 1 (CP1) that expresses the *Pse*CAST components (QCascade) promoting DNA target

binding (*43*). The selection requires targeted insertion of a transposon-encoded promoter sequence, provided by complementary plasmid 2 (CP2), upstream of a promoterless *gIII* on the AP. An SP encoding an active TnsABC variant supports promoter transposition from CP2 to AP, activating *gIII* expression and propagation of that SP. To increase selection stringency throughout evolution, we developed CP2 constructs with progressively weaker promoter strengths, requiring more integration events into the multicopy AP to trigger sufficient *gIII* expression for SP propagation before dilution out of the lagoon.

Despite high integration activity in *E. coli* (*43*), wild-type TnsABC did not support SP propagation, even on the least-stringent selection circuit (fig. S1A). This finding suggested that, under the conditions tested, wild-type TnsABC catalysis may be too slow to activate SP propagation, which requires *gIII* activation within minutes to hours of infection (*47*). Overnight incubation of TnsABC SP with host *E. coli* yielded low but detectable RNA-guided CP2 transposon integration at the AP (0.0036%) (fig. S1B), verifying that the PACE circuit can be triggered, albeit weakly, by wild-type *Pse*CAST. Luciferase reporter assays indicated that CP2 transposon integration at the AP is sufficient to activate downstream gene expression (fig. S1C), and overnight propagation assays demonstrated that *Pse*CAST expression does not interfere with phage propagation (fig. S1D). Collectively, these findings suggested that this CAST PACE selection (circuit 1.0, fig. S2A) can link the integration activity of an SP-encoded TnsABC to SP propagation if kinetically enhanced TnsABC variants enable transposition on a timescale relevant for phage replication.

Evolution of TnsABC

We initiated evolution of wild-type TnsABC using PANCE (Fig. 2A and fig. S3A), a less-stringent alternative to PACE in which dilution with fresh host cells occurs serially after overnight phage propagation (*47*). To allow weakly active SP variants to accumulate new mutations in the absence of selection, we alternated passages on the selection *E. coli* strain with passages on a "drift strain" that provides CAST-independent *gIII* expression, allowing recovery and further diversification of surviving genes (*47*). After 13 passages on host cells (PANCE N1), pooled SPs demonstrated ~10^6-fold improved overnight propagation on the selection strain and 320-fold improved integration at the AP (Fig. 2B). These data indicated that PANCE successfully linked SP propagation with the integration activity of evolving TnsABC variants.

SP from N1 propagated at levels sufficient for PACE, thus we initiated PACE (P1) with N1-derived SP (Fig. 2A and fig. S3B). After 48 hours of PACE, all evolving P1 populations were dominated by "cheating" SPs that propagated independently of TnsABC activity by acquiring a copy of *gIII*. Sequencing revealed that P1 SPs obtained *gIII* through aberrant integration of the entire posttransposition AP vector into the SP (fig. S4 and supplementary text). To reduce the risk of undesired *gIII* acquisition in future evolutions, we developed PACE circuit 1.1, which uses a split-intein pIII (*56*) with each half encoded on either the AP or CP1 (fig. S2B). With this design, full-length *gIII* acquisition by the SP would require integration or recombination of both the AP and CP1 into the same SP genome, which is unlikely.

With two integration events, instead of one, now driving full-length *gIII* expression within circuit 1.1, P1-derived SPs exhibited reduced overnight propagation compared with circuit 1.0 (fig. S3, E and F). Therefore, we performed PANCE (N2) using circuit 1.1, seeding lagoons with clonal Δ*gIII* SP from P1 (Fig. 2A and fig. S3C). After 20 passages of alternating selection and drift, the N2 SP pool still showed insufficient propagation for PACE (fig. S3F). To amplify signal from integration events, we developed circuit 1.2, which contains a modified CP1 that links integration to T7 RNA polymerase expression and places the N-terminal *gIII* half under the control of a T7 promoter (fig. S2C). Using circuit 1.2, we initiated PACE (P2), seeding lagoons with an N2 SP pool and evolving for 144 hours (Fig. 2A and fig. S3D).

SP variants from N1, P1, N2, and P2 evolution experiments exhibited increasing levels of overnight propagation and integration at the AP, indicating that PACE successfully enriched active TnsABC variants (Fig. 2B and fig. S2, E to G). Evolved variants contained diverse mutations across TnsA, TnsB, and TnsC, with generally little mutational convergence across independently evolving SP populations (table S1). Taken together, these findings indicated that PACE explored multiple trajectories for increasing TnsABC-mediated integration efficiency.

Characterization of evolved TnsABC variants

We evaluated evolved TnsABC variants in HEK293T cells, focusing on best-performing variants (through N2) and representative P2 variants (Fig. 2, C and D). Unless otherwise noted, all human-cell integration assays assessed 1-kb transposon integration efficiencies without ClpX supplementation, quantified through droplet digital polymerase chain reaction (ddPCR) (*44*) (figs. S5 to S7). Evolution through N2 substantially improved integration at two endogenous genomic loci, increasing from an average 0.062% for wild-type *Pse*CAST to an average 3.6% for the best-performing TnsABC variant, N2-1 (Fig. 2D). However, whereas the P2 SP pool supported the highest overnight propagation and AP integration (Fig. 2B), P2 TnsABC variants were substantially less active than the N2-1 variant in HEK293T cells (Fig. 2D). These findings suggested that TnsABC variants evolved fitness gains in *E. coli* during P2 that did not result in higher human-cell activity.

To better understand the disconnect between PACE fitness and human-cell integration activity, we evaluated the individual contributions of evolved TnsAB [the heteromeric transposase (*71*)] and TnsC [an AAA+ adenosine triphosphatase (ATPase) regulator of transposition (*72*, *73*)] to SP propagation in PACE and DNA integration in HEK293T cells (Fig. 2, E and F). Whereas P2-derived TnsAB and TnsC synergized to increase SP propagation (Fig. 2E), P2-derived TnsC variants reduced integration efficiencies in HEK293Ts on average by 2.8-fold compared with efficiencies mediated by wild-type TnsC (Fig. 2F). These data suggested that TnsC acquired mutations during P2 that decreased human-cell integration activity despite improving SP fitness in PACE.

Reversion analysis of P2 TnsC variants identified D44G (D44→G), D44N, and N316D, three highly conserved mutations among P2 variants (table S1), as the source of reduced activity in human cells (fig. S8, A and B). Overnight propagation assays confirmed that these mutations benefited SP propagation (fig. S8C), suggesting that TnsC-mediated determinants of *Pse*CAST activity differ between *E. coli* and human cells. According to an AlphaFold3-predicted (*74*) TnsC model, D44 is near the ATP-binding pocket, and N316 lies at the interface between adjacent TnsC monomers near the target DNA (fig. S8D). Current models of type I-F CAST mechanism (*39*, *73*) suggest that ATP binding and TnsC oligomerization are necessary for recruitment to the QCascade-bound target site. Because the DNA target search space in *E. coli* is much smaller than in human cells (*75*), we speculate that PACE optimized TnsC for improved target engagement in *E. coli* through mutations that did not benefit activity in human cells.

Evolution of TnsAB

To more effectively evolve variants that are active in human cells, we developed PACE circuit 2.0, which encodes TnsC on CP1 instead of the SP, thereby restricting evolution to TnsAB (fig. S2D). In addition, we simplified the circuit by removing split-intein *gIII* and instead encoding full-length *gIII* on an enlarged AP (10 kb). Aberrant AP recombination into the SP would yield a phage genome exceeding the M13 phage-packaging capacity (*47*), thus reducing the risk of cheating through *gIII* acquisition. During circuit 2.0 design, we reported that the transposon left ends of type I-F CASTs contain a conserved binding site for bacterial integration host factor (IHF) (*76*). IHF promotes transposition activity of some type I-F CASTs in *E. coli*, including *Pse*CAST to a weak extent (*76*). Thus, to prevent the evolution of IHF-dependent fitness, which would not translate to human cells, we mutated the IHF binding site in the transposon left end in CP2 (fig. S2D).

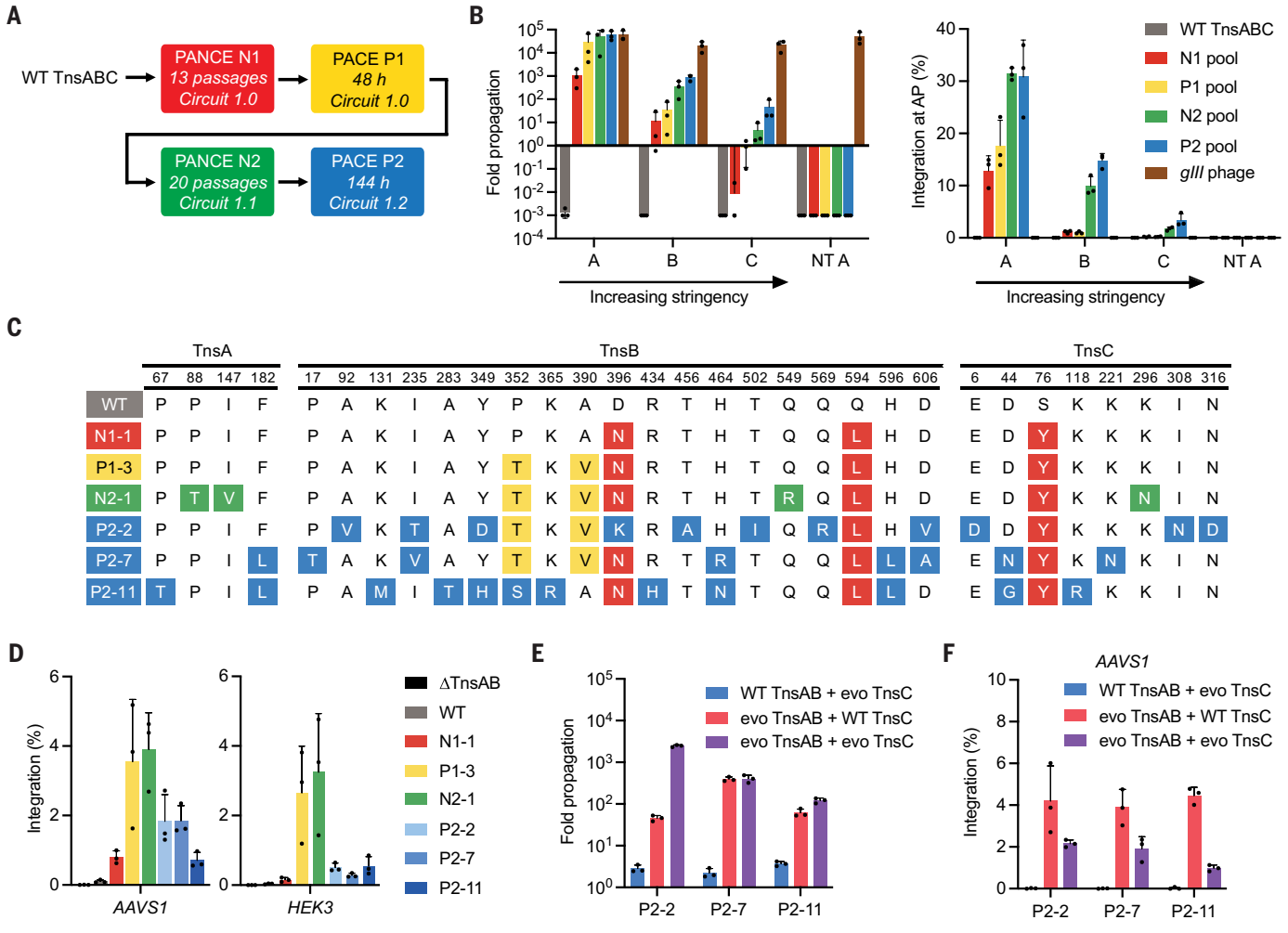


Fig. 2. Continuous evolution of TnsABC. (**A**) Summary of TnsABC evolution campaign. Whether evolution segments were conducted using PANCE or PACE is specified, with PANCE passages or PACE hours indicated. Circuit architectures are described in fig. S2, A to C. (**B**) Overnight phage propagation assays with WT TnsABC SP, pooled evolved SPs from each evolution segment, and *gIII*-expressing phage (positive control for propagation). The *x* axes indicate host *E. coli* variants encoding circuit 1.0. Host A was used for PANCE N1. Hosts B and C are of increased selection stringency, manipulated by reducing the promoter strength in the transposon on CP2 and reducing the ribosome binding site strength upstream of *gIII* on the AP. Host NT A is host A with a nontargeting crRNA. The left graph shows phage propagation levels (output phage titer divided by input titer). The right graph shows transposon integration efficiencies at the AP target site in *E. coli* after overnight propagation, measured by quantitative PCR. Promoter and ribosome binding site strengths for each host are shown in table S9. (**C**) Genotypes of a subset of evolved TnsABC variants. Variants N1-1, P1-3, and N2-1 showed the highest integration activity among the variants emerging from their respective PANCE or PACE experiments at two tested genomic sites in HEK293T cells (fig. S7). Variants P2-2, P2-7, and P2-11 are representative of the genotypes that emerged from P2 (table S1). (**D**) One-kilobase transposon integration at two genomic loci in HEK293T cells, using WT and evolved TnsABC variants specified in (C). (**E** and **F**) Assessing the contributions of P2-derived TnsAB and TnsC subunits to overnight phage propagation levels on P2 host *E. coli* (E) and 1-kb transposon integration efficiency in HEK293T cells (F). Data in (B) and in (D) to (F) are shown as mean ± SD for $n = 3$ independent biological replicates.

Selection circuit 2.0 yielded poor SP propagation (fig. S9C), which was likely the result of moving TnsC from a high-copy SP to a low-copy CP1. We therefore evolved TnsAB variants P1-3 and N2-1, which exhibited high activity in HEK293T cells, using selection circuit 2.0 in PANCE (N3) for 25 passages, alternating selection with drift through passage 16 (Fig. 3A and fig. S9A). After N3, we seeded PACE (P3) with SPs encoding N3 TnsAB variants and evolved for 140 hours (Fig. 3A and fig. S9, B and D), using a modified circuit 2.1 architecture (figs. S2E and S10 and supplementary text).

We isolated and characterized many TnsAB variants emerging from N3 and P3 (table S1 and fig. S11, A and B). In contrast to previous TnsABC evolutions (fig. S7), most evolved TnsAB variants showed improved editing in HEK293T cells (fig. S11, A and B). Mutations in TnsB alone were sufficient to achieve maximum editing levels from the top-performing variants (fig. S11C), suggesting that TnsB-related activities, which include transposon end binding and transesterification catalysis (*72*), represent key bottlenecks limiting *Pse*CAST activity in human cells.

Evolution of TnsB

Given the above findings, we restricted evolution to TnsB by developing PACE circuit 3.0, which encodes TnsA on CP1 instead of the SP (fig. S2F). We recently demonstrated that bacterial protein ClpX enhances type I-F CAST activity in HEK293T cells, although with considerable cytotoxicity (*44*) (fig. S12). To prevent evolution of ClpX-dependent fitness in PACE, we developed a host *E. coli* strain lacking endogenous *clpX*. When we used circuit 3.0, this strain reduced overnight propagation of TnsB-encoding SP by ~200-fold, whereas propagation of a *gIII*-encoding SP was unaffected (fig. S13A). These data implicated ClpX in *E. coli*-based *Pse*CAST activity and revealed an altered selection

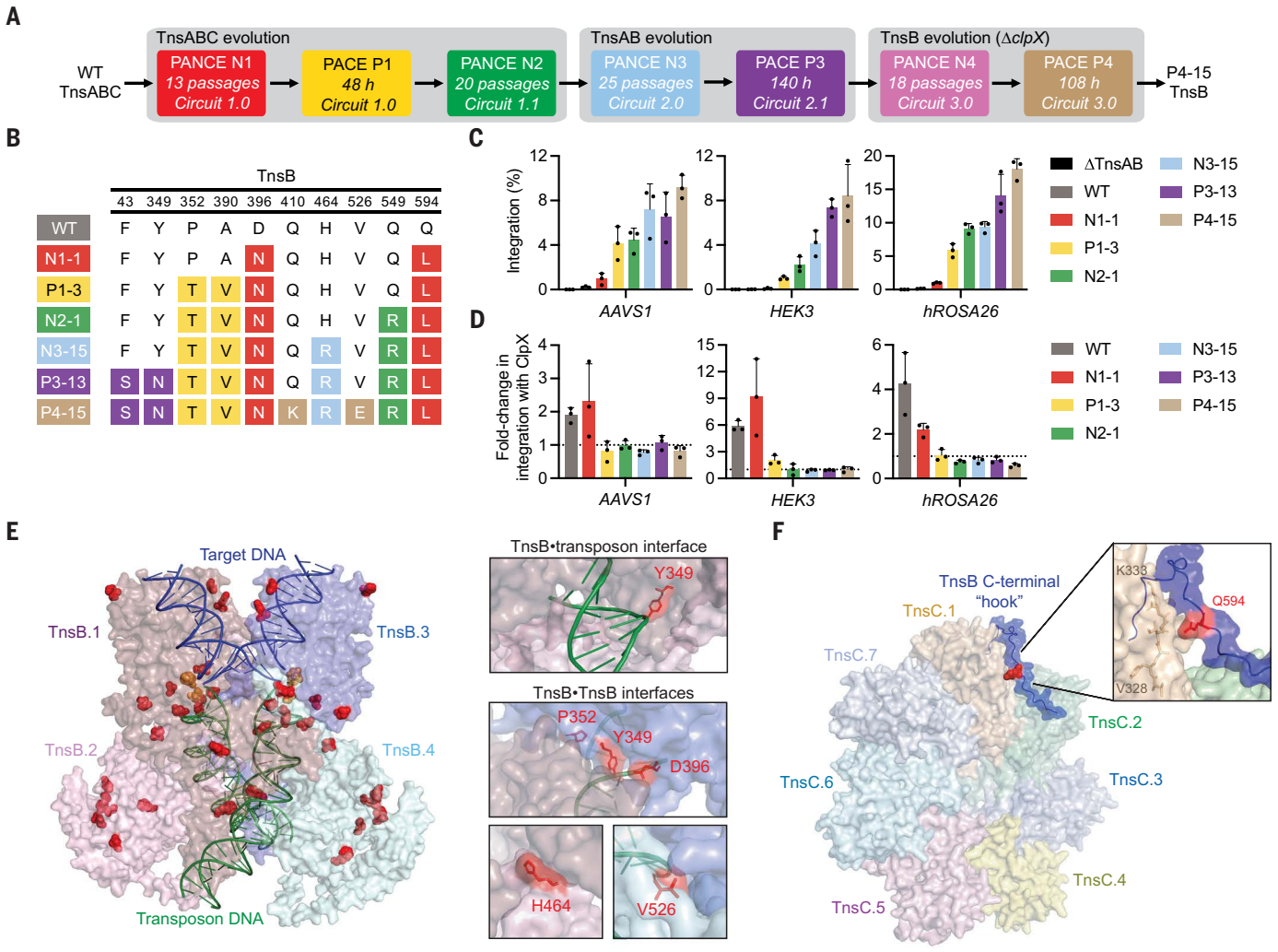


Fig. 3. TnsAB- and TnsB-focused evolution generate transposase variants that support robust integration in human cells. (**A**) Summary of the evolution campaign that yielded the evolved TnsB variant, P4-15, with the highest activity in HEK293T cells. Whether evolution segments were conducted in PANCE or PACE is specified, with PANCE passages or PACE hours indicated. Circuit architectures are described in fig. S2. (**B**) Genotypes of top-performing TnsB variants from each evolution segment. (**C**) One-kilobase transposon integration in HEK293T cells at two genomic sites by TnsB variants shown in (B). (**D**) Fold change in integration efficiencies upon cotransfection with a plasmid expressing *E. coli* ClpX. The dotted line represents no change upon ClpX expression. (**E**) Mutated residues in the P4-15 TnsB variant mapped onto an AlphaFold3-predicted structure of a *Pse*TnsAB tetramer complexed with a DNA substrate that mimics the product of TnsB transesterification. TnsA structures are omitted for clarity. The full TnsAB complex and alignment to a recently determined TnsAB structure from a type I-B CAST (*77*) are shown in fig. S16. Each transposon end (green) contains one full TnsB binding site that is joined to the 5′ end of target DNA (blue). Low-confidence unstructured C-termini of TnsB monomers [containing residues with predicted local distance difference test (pLDDT) < 70] are not shown. (Left) All mutated P4-15 residues are shown in red, with the catalytic metal-coordinating DDE residues in TnsB.1 and TnsB.3 shown in orange. (Top right) Mutated Y349 residue predicted to contact transposon DNA. (Bottom right) Multiple predicted TnsB•TnsB interfaces that contain mutated residues. (**F**) The mutated Q594 residue (red) in the P4-15 TnsB variant mapped onto an AlphaFold3-predicted structure of the *Pse*TnsB C-terminal "hook" domain in complex with a *Pse*TnsC heptamer. Data in (C) and (D) are shown as mean ± SD for $n = 3$ independent biological replicates. AlphaFold3-predicted structures in (E) and (F) are available on Zenodo (*124*). Single-letter abbreviations for the amino acid residues are as follows: A, Ala; D, Asp; E, Glu; G, Gly; H, His; K, Lys; L, Leu; N, Asn; P, Pro; Q, Gln; R, Arg; T, Thr; V, Val; and Y, Tyr.

pressure when evolving TnsB in a Δ*clpX* host, leading us to perform all evolution experiments with circuit 3.0 in this Δ*clpX* host.

We focused evolution on P3-13 TnsB, the best-performing TnsB variant from P3 (Fig. 3, E and F). P3-13 TnsB SP did not propagate sufficiently for PACE (fig. S13A), thus we initiated PANCE (N4) on P3-13 TnsB, performing 18 selection passages (Fig. 3A and fig. S13B). We used the evolved N4 SP pool to seed PACE (P4), evolving for 108 hours (Fig. 3A and fig. S13, C and D). Most P4 TnsB variants showed improved activity in HEK293T cells compared with P3-13 (fig. S14A), with the best-performing variant, P4-15, averaging 12% integration efficiency across three genomic loci (Fig. 3, B and C). Although TnsAB and TnsB evolution campaigns used a transposon left end with a mutated IHF binding site, representative P4 TnsB variants performed similarly with wild-type or mutant left ends in HEK293T cells (fig. S14B). Additionally, P4-15 TnsB, despite evolving as an unfused peptide in PACE, still mediated the highest editing in HEK293T cells when fused to TnsA through a bipartite nuclear localization signal (NLS) linker (fig. S10C), previously found to be the optimal configuration for wild-type TnsB in human cells (*44*).

Evolving P4-15 TnsB in higher-stringency host cells with reduced CP2 promoter strength failed to improve human-cell integration activity, as did introducing mutations from other highly active P4 TnsB variants into P4-15 TnsB (fig. S14). This plateau may indicate that the evolved P4-15 TnsB variant no longer bottlenecks integration efficiency under the conditions tested or that new selection pressures or evolutionary

trajectories need to be explored for TnsB PACE to continue improving integration activity. Overall, phage encoding P4-15 TnsB experienced a total 10^{322}-fold dilution over 76 PANCE passages and 296 hours of PACE, corresponding to hundreds of evolutionary generations.

Characterization of evolved TnsB variants

We next performed in-depth characterization of P4-15, the most promising evolved TnsB variant. First, we assessed whether ClpX affected integration mediated by P4-15 (Fig. 3D). ClpX is hypothesized to facilitate posttransposition complex (PTC) disassembly, enabling endogenous DNA repair machinery to access the 5-nucleotide (nt) gaps generated by staggered TnsB-catalyzed transesterifications (*44*, *71*, *72*). Whereas ClpX enhanced wild-type TnsB integration activity on average by 4.0-fold across three genomic sites in HEK293T cells, ClpX had no impact on P4-15–mediated editing at any genomic site tested, with evolutionary precursors of P4-15 also exhibiting reduced ClpX reliance (Fig. 3D). Notably, ClpX independence emerged before selection on the Δ*clpX E. coli* host (Fig. 3D), suggesting that early evolution experiments enriched variants with reduced ClpX dependence. In CAST PACE, PTC disassembly is required for SP propagation because RNA polymerase must traverse the repaired 5-nt gap to transcribe *gIII*. We hypothesize that CAST PACE enriched TnsB variants that could more rapidly disassemble after integration to more efficiently activate *gIII*, thus making ClpX less essential for integration activity.

To investigate how PACE improved TnsB activity, we mapped mutated residues in P4-15 onto two AlphaFold3-predicted (*74*) structure models: a TnsAB strand-transfer complex (*77*) (Fig. 3E and fig. S16) and a TnsB C-terminal "hook" domain in complex with a TnsC heptamer (*77*, *78*) (Fig. 3F). According to *E. coli* Tn7 biochemistry (*71*, *72*), TnsB performs multiple functions in the CAST transposition cycle, including complexing with TnsA, binding to transposon ends, binding to the target-bound TnsC, catalyzing DNA cleavage and transesterification reactions, and undergoing conformational rearrangements to allow 5-nt gap fill-in. Evolved mutations span multiple TnsB domains (fig. S16C) and predicted interfaces, including TnsB•transposon end (Y349N), TnsB•TnsB (Y349N, P352T, D396N, H464R, and V526E), and TnsB•TnsC (Q594L) (Fig. 3, E and F), suggesting that PACE evolved multiple TnsB functionalities to improve integration efficiency in human cells.

Reversion analysis of each mutation in P4-15 revealed that all 10 mutations increase activity in HEK293T cells (fig. S17A). Integration efficiency of each revertant was unchanged upon ClpX addition (fig. S17B). When installed individually into wild-type TnsB, all mutations improved efficiencies (though several only to a modest extent) except A390V (fig. S17C), which may act through epistasis. P352T and D396N, acquired early in evolution (Fig. 3B) and predicted to lie at a TnsB•TnsB interface (Fig. 3E), enabled the highest integration efficiencies among the evolved mutations when tested individually (fig. S17C). No individual mutation conferred ClpX independence (fig. S17D).

Because type I-F CASTs have yet to be reconstituted biochemically, we performed cell-based assays to illuminate the properties of evolved TnsB. We assessed transposon-end binding by P4-15, evolutionary precursors of P4-15, and P4-15 single-mutation revertants through an established transcriptional activation assay (*44*) in HEK293T cells (fig. S18A). P4-15 TnsB exhibited 3.2-fold improved reporter activation compared with wild-type TnsB, with the largest increase resulting from mutations in P1-3 (fig. S18B). Accordingly, reverting A390V—a mutation descending from P1-3—in P4-15 greatly reduced activity (fig. S18B). Residue A390 is not proximal to transposon DNA in the AlphaFold3-predicted structure (*74*) (fig. S18C), nor is it within DNA binding domains (fig. S16C), suggesting a long-distance mechanism by which A390V improves activity. Western blots revealed that evolved TnsB variants did not possess greater soluble expression than wild-type TnsB in HEK293T cells (fig. S18D), suggesting that the elevated reporter signal for the evolved TnsB variants in the transcriptional activation assay indeed resulted from enhanced TnsB•DNA binding.

We also monitored transposition by P4-15 TnsB and its evolutionary precursors through an *E. coli* liquid culture assay, in which integration results in luciferase expression (fig. S19A). Evolved variants exhibited faster apparent rates of luciferase expression (fig. S19B), suggesting that PACE indeed selected for TnsB variants with faster transposition kinetics. These results also demonstrate that evolved TnsB variants have improved activity in *E. coli*, in addition to human cells.

Collectively, these data suggest that PACE optimized diverse TnsB interactions with itself and other CAST components to improve integration activity. These findings highlight the advantage of using directed evolution to improve CAST activity, here identifying 10 activity-enhancing mutations that would be very difficult to deduce purely through rational protein engineering.

Development of evoCAST from evolved and engineered components

To further improve activity in human cells, we combined P4-15 TnsB with evolved and engineered variants of non-TnsB *Pse*CAST components that enhanced integration. After evaluating evolved TnsA and TnsC variants in combination with P4-15 TnsB (fig. S20, A and B), we identified a TnsABC combination averaging 1.3-fold improved integration across four genomic sites in HEK293T cells compared with P4-15 TnsB with wild-type TnsA and TnsC (fig. S20C). Improvements mediated by evolved TnsAB were independent of TnsAB stoichiometry (fig. S20D), although the extent to which evolved TnsC improved efficiencies depended on the amount of TnsC delivered (fig. S20E and supplementary text).

Previous work revealed that *Pse*CAST's QCascade complex exhibits weaker DNA binding activity than other type I-F CAST QCascades (*44*, *78*). We attempted to evolve DNA-targeting components in PACE, encoding QCascade on the SP and TnsABC on CP1, but failed to isolate QCascade variants that increased integration efficiency in HEK293T cells. This failure may have arisen from the many differences between DNA targeting in *E. coli* and human cells, including chromatinization, DNA supercoiling, and target search space. As an alternative, we engineered a QCascade module that enhanced integration activity, incorporating (i) a Cas7 with a neutral DNA-contacting residue mutated to lysine, (ii) a Cas8 containing an engineered PAM-interacting domain previously found to improve wild-type *Pse*CAST activity in HEK293T cells (*78*), and (iii) an additional bipartite NLS at the N-termini of TniQ, Cas6, and Cas8 (fig. S21 and tables S2 to S4).

By combining a PACE-evolved TnsABC with a rationally engineered QCascade, we developed a CAST system (evoCAST) optimized for human cells (Fig. 4A). Across four genomic sites in HEK293T cells, evoCAST averaged 19% integration, representing an average 1.2-fold improvement over P4-15 TnsB with unoptimized non-TnsB *Pse*CAST components and an average 540-fold improvement over wild-type *Pse*CAST (Fig. 4B). EvoCAST integrated a range of DNA payload sizes up to 15 kb, the largest size tested (Fig. 4C). EvoCAST also supported both plasmid- and linear-donor DNA topologies (fig. S22). Together, the improvements made to all seven *Pse*CAST protein components establish evoCAST as a platform for targeted genomic integration of gene-sized DNA cargoes in mammalian cells.

Characterization of evoCAST integration products

Next, we examined evoCAST integration products in HEK293T cells. High-throughput sequencing (HTS) of genome-transposon junctions revealed that evoCAST retained the wild-type *Pse*CAST insertion site preference, integrating ~49 bp downstream of the RNA-complementary target sequence (Fig. 4D). In contrast to nuclease-mediated end-joining (*27*) or HDR (*2*–*4*), evoCAST mediated efficient DNA integration without detected indel formation at unintegrated loci (Fig. 4E). HTS revealed low levels (<3%) of substitution mutations within the 5-bp target-site duplication (TSD) for evoCAST integration products (fig. S23), which may have arisen from host repair of the 5-nt gaps generated by offset TnsB transesterifications (*71*). To assess the orientation of evoCAST integration products, we performed ddPCR with orientation-specific probes. Across four genomic loci tested, evoCAST exhibited a strong

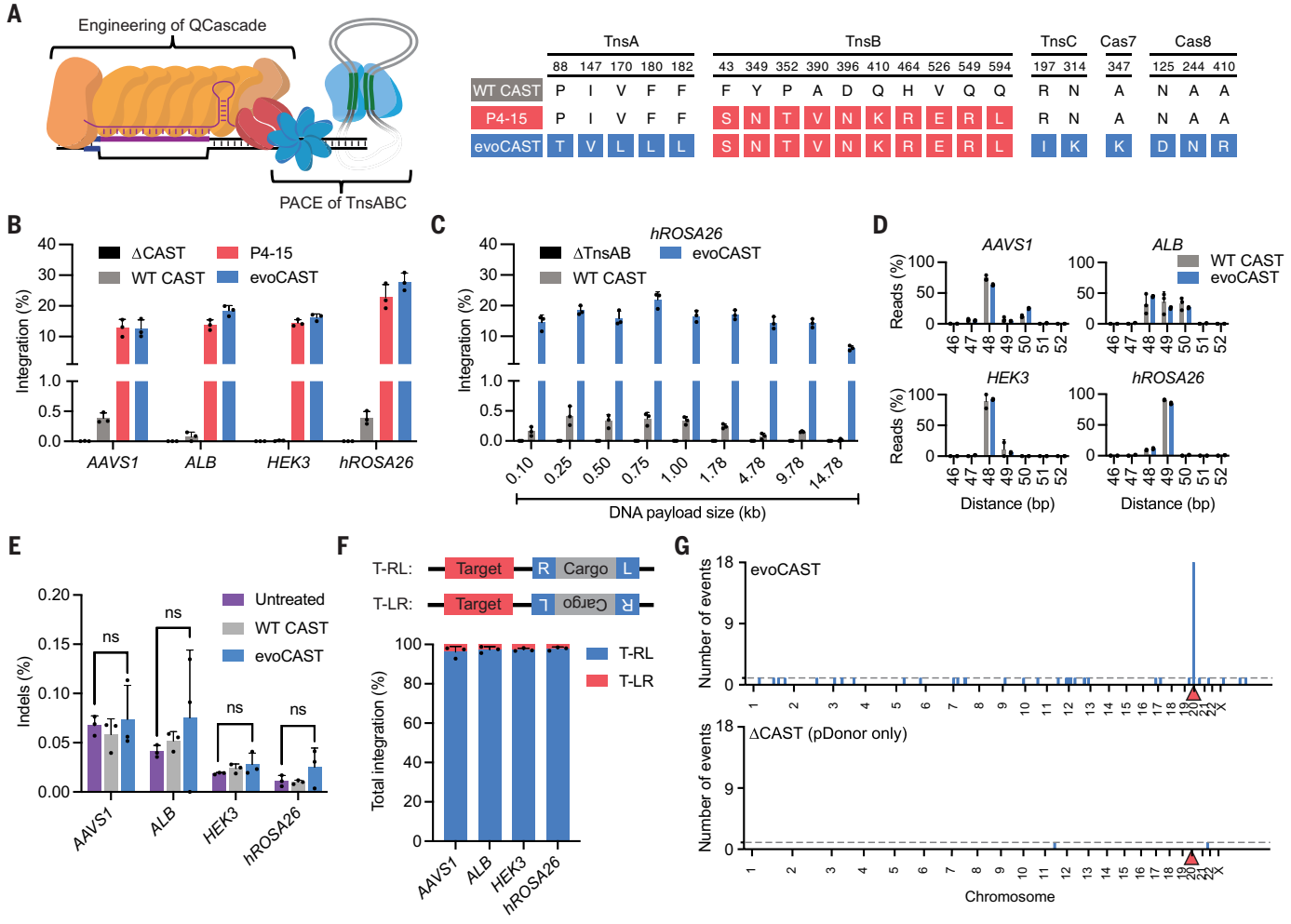


Fig. 4. Development and characterization of evoCAST. (**A**) Schematic (left) and genotypes (right) of P4-15 TnsB and evoCAST components. EvoCAST also contains optimized NLS architectures for Cas6, Cas8, and TniQ. (**B**) One-kilobase transposon integration efficiencies by evoCAST compared with P4-15 TnsB and WT *Pse*CAST at four genomic sites in HEK293T cells. (**C**) Integration of varying DNA payload sizes (measured as the distance between the 3′ end of the transposon right end and 5′ end of the transposon left end) by WT *Pse*CAST and evoCAST in HEK293T cells. Donor DNA transfected was normalized by mass. (**D**) HTS analysis of the distance between the 3′ end of the target site and 5′ end of the transposon integration site for WT *Pse*CAST and evoCAST across four genomic sites in HEK293T cells. (**E**) Comparison of indel formation across untreated cells, WT *Pse*CAST, and evoCAST at four genomic sites in HEK293T cells. Indels were quantified across a 40-bp window centered at the predicted insertion site for all unintegrated reads (Materials and methods). An unpaired, two-sided *t* test was performed to determine statistical significance, with "ns" (not significant) indicating a *P* value > 0.05. (**F**) Relative frequencies of integration in the T-RL or T-LR orientation for evoCAST across four genomic sites in HEK293T cells, determined by ddPCR, using probes specific to either T-RL or T-LR integration events. (**G**) Genome-wide integration events for evoCAST (top) and a negative control (bottom) in which only pDonor was transfected, detected through a modified UDiTaS workflow (*81*). Integration events were measured by the number of distinctive molecular identifiers identified at a single integration site (Materials and methods). The on-target genomic site (*AAVS1*) is indicated with a red triangle. The dotted line corresponds to a single detected integration event. Shown is one of two replicates; both replicates are shown in table S5. Data in (B) to (F) are shown as mean ± SD for $n = 3$ independent biological replicates.

preference for integrating the right transposon end proximal to the Cascade-targeted site (Fig. 4F), similar to the preference of wild-type *Pse*CAST (*43*). Long-read sequencing of insertion product amplicons (*79*) indicated that both evoCAST and wild-type *Pse*CAST products were predominantly simple insertions, not cointegrates—undesired by-products containing two cargo DNA copies and the donor vector backbone that are caused by an absence of TnsA endonuclease activity (*41*, *79*, *80*) (fig. S24). Collectively, these results demonstrate evoCAST's high product purity, characterized by single–base pair precision, no detected indels, near unidirectionality, and predominantly simple insertions over cointegrates.

Characterization of off-target evoCAST integration

To evaluate evoCAST's genome-wide specificity in human cells, we used a modified unidirectional targeted sequencing (UDiTaS) approach (*81*) (fig. S25A), which we previously applied to recombinases in bacteria and human cells (*37*, *82*). Although type I-F CASTs show high specificity in *E. coli* (*39*, *43*), we could not assess the specificity of wild-type *Pse*CAST in human cells owing to very low integration efficiency.

We assessed the specificity of evoCAST targeting *AAVS1* in HEK293T cells after 1 week of incubation with plasmid expression vectors. Although on-target integration was by far the most common outcome, genomic off-target integration events were detected. Each detected off-target integration was observed only once, suggesting that off-target integration occurred primarily as distinct single events (Fig. 4G and table S5). Across two replicates, evoCAST averaged 36% on-target integration (table S6), which is similar to the fraction of on-target editing events calculated from published data for Cas9 nuclease (average 47%) (*83*) and programmable gene integration with

eePASSIGE (0.51 to 38%, depending on the attachment site used in the donor DNA) (*37*) (table S6).

We next investigated the source of evoCAST off-target formation. Each off-target integration site lacked homology to the *AAVS1* target site and was not reproducible across replicates (table S5). Additionally, off-target integration required TnsC but not QCascade (fig. S25B). Taken together, these data suggest that off-target integration is CRISPR-independent and arises from aberrantly bound TnsABC complexes. Off-target events persisted when using evolved TnsB with wild-type TnsA and TnsC (fig. S25C), suggesting that the evolved TnsB may be responsible for integration at off-target substrates transiently engaged by the TnsABC complex. Off-target sites were generally in regions of open chromatin (fig. S25D), which are likely to be more accessible substrates for TnsABC. EvoCAST off-target sites did not contain disproportionately high AT (adenine-thymine) content (fig. S25E), suggesting a different mechanism of off-target formation than previously reported for type V-K CAST (*84*).

Additional off-target-profiling experiments offered insight into how off-target formation could impact evoCAST applications. In cell populations enriched for on-target editing, off-target integration events were not detected (fig. S25F), suggesting that on-target integration in cells is generally not accompanied by off-target integration events. UDiTaS of *E. coli* lysate from PACE experiments with P4-15 TnsB-encoding SP revealed >99% on-target integration (fig. S25G), suggesting that transient evoCAST exposure (hours during PACE versus days in HEK293T cells) maximizes the ratio of on-target:off-target integration, which is consistent with the behavior of other genome editing agents (*4*, *85*–*91*). Lastly, TnsABC-mediated off-target integration was not detected by a fluorescent reporter assay previously used to detect off-target formation by eeBxb1 recombinase with an *attP*-containing donor (*37*) (fig. S25H), which is consistent with the UDiTaS data demonstrating that evoCAST mediates lower off-target integration than this eePASSIGE configuration (table S6).

Although off-target formation by evoCAST warrants further investigation, these data indicate that the proportion of on-target versus off-target editing events for evoCAST is similar to those of other current genome editing methods. Additionally, current UDiTaS methods have limited coverage of integration events (tables S5 and 6), owing to a substantial fraction of sequencing reads aligning to unintegrated pDonor molecules (Materials and methods). Further development of off-target characterization methods for targeted gene integration may enable more precise quantification of off-target editing frequencies in mammalian cells for CASTs and other large DNA integration technologies.

Application of evoCAST at genomic target sites of therapeutic interest

We next applied evoCAST for targeted gene-sized DNA integration at therapeutically relevant human genomic sites. After optimizing CRISPR RNA (crRNA) architecture and spacer sequences (fig. S26), we assessed integration at 14 genomic loci in HEK293T cells (Fig. 5A). EvoCAST averaged 14% 1-kb transposon integration efficiency at these target sites, compared with 0.22% for wild-type *Pse*CAST (Fig. 5A). EvoCAST efficiencies were positively correlated with chromatin accessibility (fig. S27), which is similar to previous reports for other genome editing methods (*37*, *92*, *93*).

Therapeutic gene integration at *ALB* is a promising strategy for transgene expression in hepatocytes (*94*, *95*). *ALB* is highly expressed in the liver, and integrating a splice acceptor–bearing donor into intron 1 enables splicing with an exon 1–encoded secretion signal for subsequent protein secretion (*94*). This strategy is being investigated for treatment of hemophilia B, in which restoring 1% of circulating human factor IX (hFIX) levels is therapeutic (*94*, *95*). We integrated *F9* cDNA encoding the hyperactive hFIX Padua variant (*96*) into *ALB* intron 1 in a human hepatocyte cell line (HuH7) (Fig. 5B), with evoCAST achieving 5.7% integration efficiency (0.023% for wild-type *Pse*CAST) (Fig. 5E) and enabling detectable *F9* expression (which was not detected for wild-type *Pse*CAST) (fig. S28A).

Integration of a CAR at the T cell receptor α constant (*TRAC*) locus enables uniform CAR expression, enhanced T cell potency, and delayed T cell exhaustion (*97*). We assessed the efficiency of CD19 CAR integration at *TRAC*, a strategy shown to combat refractory or relapsed B cell malignancies (*98*) (Fig. 5, C and F). In HEK293T cells, evoCAST mediated 13% integration of CD19 at *TRAC,* compared with 0.061% by wild-type *Pse*CAST (Fig. 5F).

The programmability of evoCAST potentiates integration of wild-type cDNAs at sites of endogenous gene mutation or deletion that are associated with loss-of-function diseases (Fig. 5D). This strategy could treat loss-of-function diseases in an allele-agnostic manner while preserving some endogenous regulatory context. We assessed integration of wild-type cDNAs (Δexon 1, flanked by a 5′ splice acceptor and 3′ polyA signal) into intron 1 of *FANCA* (associated with Fanconi anemia), *IL2RG* (X-linked severe combined immunodeficiency), *MECP2* (Rett syndrome), and *PAH* (phenylketonuria) (Fig. 5G). EvoCAST supported substantial targeted gene insertion efficiencies of 12 to 15% at these loci, compared with 0.0092 to 0.43% by wild-type *Pse*CAST (Fig. 5G). We measured integrated transgene expression at *MECP2*—which is expressed in HEK293T cells (*99*)—through reverse transcription ddPCR using a transgene-specific probe and found that evoCAST, but not wild-type *PseCAST*, yielded detectable transgene expression (fig. S28B).

To enable broader applications of evoCAST, we engineered transposon end sequences that do not compromise integration activity but are compatible with in-frame protein tagging (fig. S29 and supplementary text). We also demonstrated that evoCAST-edited cells persist in bulk populations when using a selectable marker and that clonally integrated populations can be isolated through single-cell sorting (fig. S30 and supplementary text).

Collectively, these results demonstrate that evoCAST can be reprogrammed to integrate large, diverse DNA payloads across multiple genomic loci in human cells, enabling a range of potential applications in therapeutic science and basic research.

EvoCAST in other mammalian cell types

We next characterized evoCAST in multiple human cell types. We assessed 1-kb transposon integration at two genomic sites in HeLa and K562 cells (Fig. 5H). In HeLa cells, evoCAST averaged 4.7% editing activity, compared with 0.18% by wild-type *Pse*CAST (Fig. 5H). In K562 cells, evoCAST averaged 1.6% editing, compared with 0.038% by wild-type *Pse*CAST (Fig. 5H). The lower editing activity in these cell types compared with that in HEK293T cells may arise from their less-efficient transfection of the multiple vectors encoding CAST components along with the donor DNA. Optimizing delivery methods may further improve evoCAST efficiencies in difficult-to-transfect cell lines.

To demonstrate utility in a more therapeutically relevant cell type, we evaluated evoCAST in primary human fibroblast cells isolated from a patient with recessive dystrophic epidermolysis bullosa (RDEB), a mutationally diverse loss-of-function disease that can be treated using autologous gene-corrected fibroblasts (*100*). EvoCAST averaged 19% efficiency of 1-kb transposon integration across two genomic loci, representing a 200-fold improvement over wild-type *Pse*CAST (Fig. 5I). These data also establish that primary human cells can support evoCAST activity, especially when cultured under conditions that mitigate the toxicity of exogenous DNA (Materials and methods).

We also assessed the impact of ClpX codelivery on evoCAST efficiency at two genomic loci in HEK293T, HeLa, K562, and primary human fibroblast cells (Fig. 5J and fig. S31, A and B). As expected, evoCAST did not require ClpX for maximal editing in HEK293T cells (Fig. 5J). ClpX increased evoCAST editing in other cell types, notably enabling an average 37% editing in primary fibroblasts (fig. S31B),

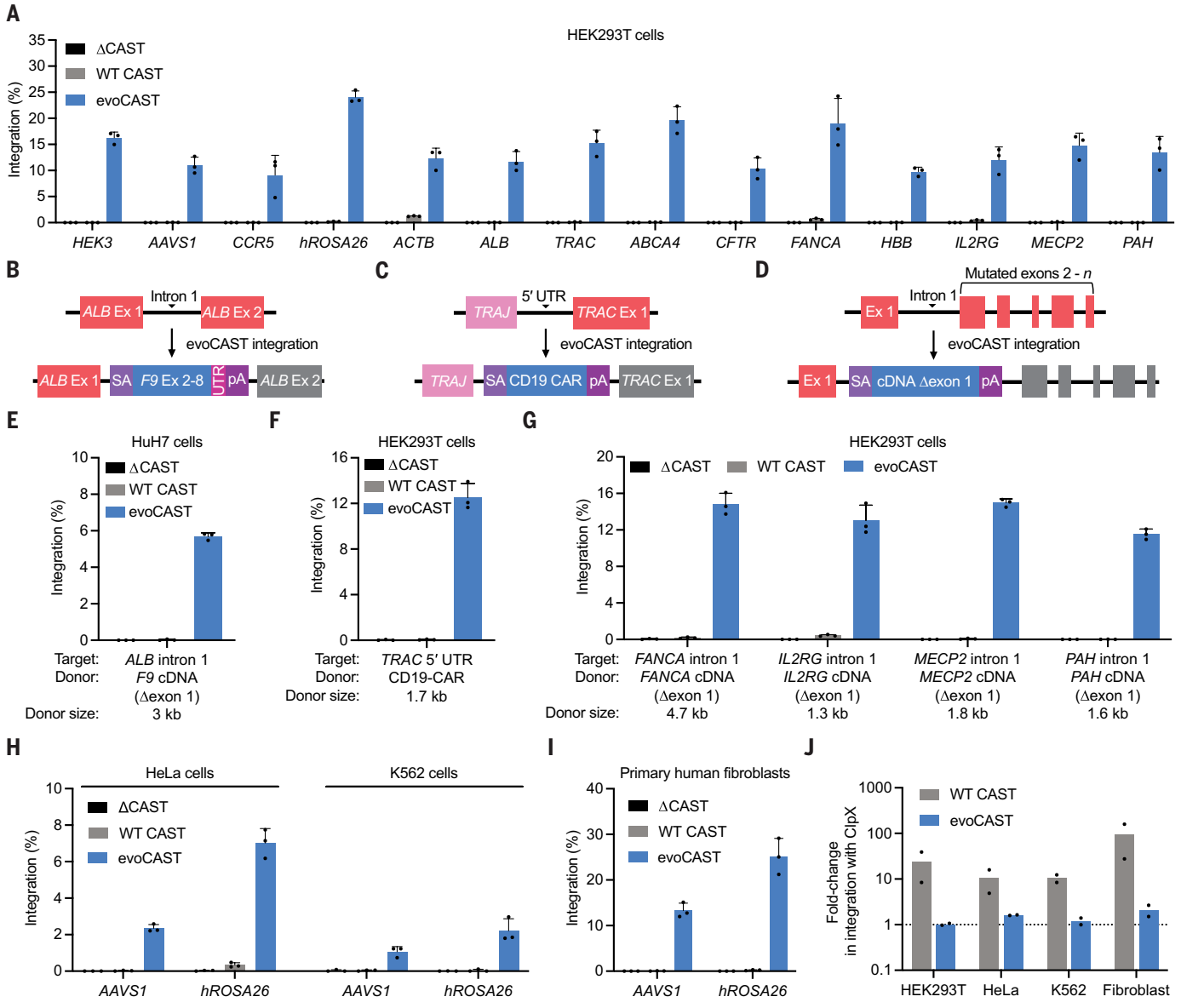


Fig. 5. evoCAST mediates efficient DNA integration at therapeutically relevant endogenous genomic loci in multiple human cell types. (**A**) One-kilobase transposon integration by WT PseCAST and evoCAST at 14 genomic loci in HEK293T cells. Each locus was targeted using a top-performing crRNA identified in fig. S26, except *HEK3*, which did not undergo crRNA spacer optimization. (**B** to **D**) Schematics depicting evoCAST applications for integrating a *F9* cDNA at *ALB* intron 1 (B), a CD19-targeted CAR at the 5′ untranslated region (UTR) of *TRAC* (C), and a cDNA encoding a healthy gene copy (Δexon 1) into intron 1 of a gene associated with pathogenic loss of function (D). (**E** to **G**) Integration by WT *Pse*CAST and evoCAST of *F9* cDNA into *ALB* intron 1 in HuH7 cells (E), CD19-CAR into the 5′ UTR of *TRAC* in HEK293T cells (F), and WT cDNAs (Δexon 1) into intron 1 of their corresponding endogenous locus in HEK293T cells (G). (**H**) One-kilobase transposon integration by WT *Pse*CAST and evoCAST at two genomic loci in HeLa and K562 cells. (**I**) One-kilobase transposon integration by WT *Pse*CAST and evoCAST at two genomic loci in primary human fibroblast cells isolated from an RDEB patient. (**J**) Fold change in integration efficiencies upon cotransfection with a plasmid expressing *E. coli* ClpX. The dotted line represents no change upon ClpX expression. Each datapoint represents the average fold change in 1-kb transposon integration efficiency at one genomic target across three independent biological replicates. Integration efficiencies used to calculate fold change are shown in fig. S32. Data in (A), (B), and (F) to (I) are shown as mean ± SD for $n = 3$ independent biological replicates.

although with increased cytotoxicity (fig. S31C). Nonetheless, evoCAST displayed substantially reduced ClpX dependence compared with wild-type *Pse*CAST across all cell types (Fig. 5J), enabling CAST-mediated editing human cells without ClpX-associated cytotoxicity.

EvoCAST compared with eePASSIGE

Lastly, we compared evoCAST with eePASSIGE (*37*), a recently described method for targeted DNA integration in human cells that does not require genomic DSBs. We assessed 2-kb DNA cargo integration at six genomic loci in HEK293T cells for both strategies (fig. S32A), although the exact target sites differed between methods because of distinct DNA targeting mechanisms. Consistent with previously reported efficiencies (*37*), eePASSIGE exhibited 1.5- to 2.9-fold-higher integration efficiency than evoCAST, although both methods typically achieved >10% integration (fig. S32A). Both evoCAST and eePASSIGE supported 10 to 40% integration efficiencies in primary human fibroblasts (*37*) (Fig. 5I and fig. S31B).

The distinct mechanisms of evoCAST and eePASSIGE result in different advantages. Although more efficient on average, the eePASSIGE method

requires attachment site installation by prime editing before DNA integration, resulting in a more heterogeneous mixture of editing outcomes than is observed with evoCAST (fig. S32, B to D). EvoCAST's high product purity makes it well-suited for applications in which by-products such as indels or unintegrated attachment site installation must be minimized. Moreover, eePASSIGE can require testing many prime editing conditions for efficient attachment site installation (*35*, *37*, *101*). By contrast, we identified high-performance evoCAST crRNAs after testing only 5 to 10 constructs per genomic locus (fig. S26, A and D).

EePASSIGE and evoCAST also differ in their compatibility with donor DNA substrates. Recombinase-mediated integration of linear DNA yields genomic DSBs, a result that is not expected with evoCAST (fig. S32E). Indeed, when using a linear donor, 36% of eePASSIGE integration products contained indels consistent with DSB formation, whereas evoCAST products contained no detected indels above background (fig. S32, F to H). EvoCAST's compatibility with linear-donor topology enables PCR amplicons to be more readily used as substrates for targeted integration. This feature is also advantageous for viral delivery modalities such as adeno-associated virus (AAV) that deliver linear DNA (*102*). Lastly, when using a circular donor DNA substrate, evoCAST integrates the desired DNA payload without flanking vector sequences. By contrast, recombinase-mediated integration installs the entire vector sequence, which can include undesired DNA elements. Taken together, these findings demonstrate the strengths of evoCAST as a platform for single-step, programmable gene integration in human cells.

Discussion

Through the development and optimization of CAST PACE, we evolved transposase variants that catalyze substantially improved (often >100-fold) genomic integration activity in human cells. Evolution in *E. coli*, the host cell for PACE, required extensive optimization because early evolution campaigns enriched some mutations that increased fitness during PACE without improving performance in human cells. Although we initially hypothesized that coevolving all three transposase complex subunits would be an effective approach to improve human-cell integration activity, we found that evolving TnsB alone was the most efficient strategy to generate CAST variants with robust activity in human cells (Fig. 3C).

The CAST PACE results suggest that *Pse*TnsB activity, which naturally evolved to support transposon mobilization in bacteria, is insufficient to enable efficient genome editing in human cells. Suboptimal transposase activity may not limit integration efficiency in *E. coli*—where *Pse*CAST enables >99% genomic integration (*43*)—owing to high transposase concentration from plasmid overexpression coupled with reaction conditions that more closely resemble the native bacterial context in which *Pse*CAST evolved. By contrast, the corresponding differences in mammalian cells may have reduced the efficiency of *Pse*TnsB-catalyzed transposition enough to become the primary bottleneck in mammalian cells.

Given that structural and biochemical characterization of the transposase complexes from type V-K (*103*, *104*) and type I-B (*77*) CASTs have substantially progressed our understanding of how these systems achieve targeted integration, we anticipate that further study of type I-F transpososomes will help elucidate how the evolved mutations across transposase subunits both improved and hindered human-cell integration activity. One important outcome of TnsB PACE is the reduced dependence on ClpX for efficient genomic integration in human cells (Figs. 3D and 5K), suggesting that CAST PACE could serve as a platform for decoupling CASTs from cytotoxic factors previously required for even low levels of editing in human cells.

The detection of off-target integration events from P4-15 TnsB (Fig. 4G) is consistent with its greatly enhanced transposition activity, which may increase integration at off-target sites transiently engaged by the transposase complex. The robust specificity of P4-15 TnsB within the PACE circuit (fig. S25G) suggests that evolved variants are not simply low-fidelity transposases; instead, off-target events detected in human cells may be a consequence of the high activity of evolved TnsB, coupled with the extended exposure of genomes to plasmid-driven transposase overexpression. Applying more transient delivery modalities, such as mRNA or ribonucleoprotein delivery, may reduce off-target integration by minimizing exposure of genomes to transposase after integration at the target site is complete, as has been observed previously with Cas9 nucleases, base editors, and prime editors (*4*, *85*–*91*). The multicomponent nature of CASTs offers additional potential strategies for improving target specificity, including engineering TnsC for reduced off-target DNA engagement, engineering TnsB so that only stably bound on-target sites promote integration, and dissecting the conformational distinctions that follow on-target versus off-target site interrogation (*84*, *105*).

Collectively, the development of evoCAST—combining advancements from both evolution and rational engineering—enabled an average 420-fold improvement over wild-type *Pse*CAST across the 14 genomic sites tested in this study (Fig. 5A). EvoCAST showed enhanced activity across diverse cell types, notably enabling up to 38% editing in unsorted primary human fibroblasts (fig. S31B), representing, to our knowledge, the highest CAST-mediated editing efficiencies at human genomic sites reported to date. EvoCAST offers distinct advantages as a genome editing platform—such as its facile reprogramming, high product purity, and avoidance of genomic DSB formation—that stem from its origins as a naturally occurring system for RNA-guided DNA insertion (*38*, *39*).

Additional research is needed to understand the determinants of optimal crRNA selection; assess the long-term stability of integrated transgene expression; further optimize integration efficiencies in diverse cell types, including nondividing cells; and develop strategies for delivering donor DNA into cell types that poorly tolerate foreign DNA (*106*). Lastly, we anticipate that lessons learned from the development of CAST PACE can be applied to other naturally occurring (*43*, *107*, *108*) and engineered (*78*, *79*) CAST systems, providing strategies to generate a suite of human cell–active CAST systems that each potentially offer distinctive advantages for targeted gene-sized DNA integration.

Materials and methods

General methods

Antibiotics were purchased from Gold Biotechnology and used at the following concentrations: streptomycin (50 μg/ml), chloramphenicol (25 μg/ml), carbenicillin (50 μg/ml), spectinomycin (50 μg/ml), tetracycline (10 μg/ml), and kanamycin (25 μg/ml). PCRs were performed using Phusion U Green Multiplex PCR Master Mix (Thermo Fisher Scientific) or Q5 Hot Start High-Fidelity 2× Master Mix (New England BioLabs) unless otherwise noted. DNA oligonucleotides, including FAM/Iowa Black FQ-labeled DNA oligonucleotides, were obtained from Integrated DNA Technologies. Human codon-optimized wild-type *Pse*CAST genes were synthesized by GenScript. All plasmids used in this study were cloned using USER, Golden Gate, or Gibson assembly methods as described previously (*9*, *55*). Plasmids were cloned into Mach1 (Thermo Fisher Scientific) chemically competent *E. coli*. Unless otherwise noted, plasmid DNA was amplified using the Illustra Templiphi 100 Amplification kit (GE Healthcare Life Sciences) prior to Sanger sequencing (Quintara Biosciences) or Nanopore sequencing (Plasmidsaurus). All plasmids for *E. coli* experiments were purified using QIAprep Spin Miniprep Kits (Qiagen), and all plasmids for mammalian cell experiments were purified using Plasmid Plus Midiprep Kits (Qiagen) or Plasmid Plus 96 Miniprep Kits (Qiagen). All isolated plasmid DNA was eluted in nuclease-free water and quantified using a NanoDrop ONE UV-Vis spectrophotometer (Thermo Fisher Scientific). A list of plasmids and selection phages (SPs) used in this work is provided in table S9. All gRNA sequences and DNA cargoes used in mammalian cell experiments are listed in tables S10 and S11, respectively. Primer sequences used for quantification, strain generation, and linear donor generation are listed in table S12. EvoCAST component sequences are listed in table S13.

Preparation and transformation of chemically competent cells

Strain S2060 (*47*) was used in all luciferase, phage propagation, and plaque assays, and in all PACE experiments, except for PACE campaigns conducted on the Δ*clpX* strain generated in this study. Chemically competent cells were prepared as described previously (*47*). Briefly, an overnight culture of bacteria was diluted 50 to 200-fold into 2×YT media (United States Biologicals) with appropriate antibiotics and grown at 37°C, shaking at 230 RPM until the culture reached an optical density (OD_{600}) of 0.4 to 0.6. Cells were centrifuged at 4°C for 5 to 10 min at 4000 *g*. The supernatant was discarded, and cell pellets were resuspended in ice-cold TSS solution (LB media supplemented with 5% v/v DMSO, 10% w/v PEG3350, and 20 mM $MgCl_2$). Resuspended cells were aliquoted, flash frozen on dry ice, and stored at −80°C until use.

To transform cells, 100 μl of competent cells thawed on ice were added to a pre-chilled mixture of plasmid (1 to 2 μl each; up to three plasmids per transformation) in 100-μl KCM solution (100 mM KCl, 30 mM $CaCl_2$, and 50 mM $MgCl_2$ in H2O) and stirred gently. The mixture was incubated on ice for >5 min, heat shocked at 42°C for 90 s, and combined with 500 μl of SOC media (New England BioLabs). Cells recovered at 37°C, shaking at 230 RPM for 1 hour. Cells were then streaked on 2×YT media + 1.5% agar (United States Biologicals) plates containing appropriate antibiotics and incubated for 16 to 18 hours at 37°C.

Bacteriophage cloning

Phage were cloned using USER assembly as previously described with minor modification (*66*). Briefly, a 25-μl USER assembly was transformed into 100 μl of chemically competent S2060 *E. coli* host cells containing pJC175e (S2208) (*47*), which enables activity-independent phage propagation. Transformed S2208 were incubated overnight at 37°C in 10-ml 2×YT media shaking at 230 RPM. The saturated culture was then centrifuged for 5 min at 4000 *g*, and the phage-containing supernatant was plaqued as described below. Individual phage plaques were grown in DRM media (United States Biologicals) for 6 to 8 hours. Following incubation, the cultures were centrifuged for 5 min at 4000 *g*, and the phage-containing supernatants were filtered through a 0.22-μm PVDF ultrafree centrifugal filter (Millipore) to remove residual bacteria. Phage were then sequenced via PCR amplicon sequencing (Quintara Biosciences), with sequence-confirmed phage stored at 4°C until use.

Plaque assay

Plaquing was performed as previously described (*66*). In brief, a saturated S2208 culture was back-diluted 50- to 100-fold into DRM containing carbenicillin. Cells were grown at 37°C shaking at 230 RPM to an OD_{600} of 0.4 to 1.0, at which point they were placed on ice during preparation of phage. Phage stocks were serially diluted in water by a factor of 10, up to 10^6-fold. 10 μl of phage stock and dilutions (typically 10^2-, 10^4-, and 10^6-fold dilutions) were combined with 100 μl of mid-log S2208 cells in 2-ml library tubes (VWR international). 1 ml of warm top agar (3:2 mixture of 2×YT medium and molten 2×YT medium agar [1.5%, resulting in a 0.6% agar final concentration], stored at 55°C until use) was added to the phage/bacteria solution, mixed once by pipetting, and then immediately plated onto one quadrant of a 2×YT medium 1.5% agar plate containing no antibiotics and 0.08% Bluo-gal (Gold Biotechnology). The plates were left to sit for 2 min undisturbed at room temperature, and then plates were incubated, without inverting, at 37°C overnight. Phage titers were determined by quantifying blue plaques. For higher-throughput plaquing, the reagents were adjusted for the wells of a 12-well plate as follows: 450 μl of top agar, 10 μl of phage, and 100 μl of cells.

Overnight propagation assay

For each replicate, a single colony of a *E. coli* host strain was picked and grown overnight at 37°C with shaking at 230 RPM in DRM and appropriate antibiotics. Saturated cultures were back-diluted 50-fold into DRM with appropriate antibiotics and grown for ~2 hours at 37°C with shaking at 230 RPM until they reached an OD_{600} of ~0.4. 1 ml of culture was then added to a 96-well deep well plate (Axygen) and infected with 1E5 total phage. This mixture was then grown overnight at 37°C with shaking at 230 RPM, and then centrifuged for 10 min at 3400 *g*. Phage-containing supernatant was then collected and plaqued to determine the total number of output phage. Fold propagation was calculated by dividing the number of output phage by the number of input phage.

qPCR quantification of transposition efficiency in *E. coli*

Quantification was performed as previously described (*39*) with modification. Two primer pairs were designed: one pair specific to the AP-transposon junction generated by T-RL integration at the AP target site, and the second pair specific to the AP backbone (primer sequences listed in table S12). Input *E. coli* lysate was prepared by resuspending a cell pellet from an overnight propagation assay in 1 ml water and incubating 50 μl of this solution at 95°C for 10 min. Standards were generated by mixing mock-T-RL-integrated AP plasmid and unintegrated AP plasmid at varying ratios, corresponding to integration efficiencies spanning 0.0064 to 100%. qPCR reactions contained 10 μl Q5 Hot Start High-Fidelity 2× Master Mix (New England BioLabs), 0.50 μM of forward and reverse primer, 0.2 μl 100× SYBR Green (Invitrogen), 4 μl of standard or 100-fold diluted lysate, and nuclease-free water to 20 μl total volume. qPCR was run on a BioRad CFX96 Real Time system with the following cycling conditions: 98°C for 2 min; 40 cycles of 98°C for 10 s, 60°C for 20 s, and 72°C for 15 s. Each sample was analyzed in two parallel reactions: one reaction with the primer pair specific to T-RL integration, the second reaction with the primer pair specific to the AP backbone. Transposition efficiency was calculated using a linear regression generated by the ΔC_q values (C_q difference between the two parallel qPCR reactions) of the standards containing known integration efficiencies.

Luciferase assay

S2060 cells (*47*) were transformed with necessary plasmids. For endpoint luciferase assays (fig. S1C), saturated overnight cultures of single colonies were diluted 250-fold into DRM media with appropriate antibiotics and grown for ~3 hours at 37°C with shaking at 230 RPM. 100 μl of cells were transferred to a 96-well black-walled clear-bottomed plate (Costar), then 600-nm absorbance and luminescence were read using a plate reader (Tecan). For time-course luciferase assays (fig. S19), saturated overnight cultures of single colonies were diluted 250-fold into 2×YT media with appropriate antibiotics and grown for ~2 hours at 37°C with shaking at 230 RPM to an OD_{600} of ~0.2. 100 μl of cells were then combined with 10 mM arabinose and transferred to a 96-well black-walled clear-bottomed plate (Costar). 600-nm absorbance and luminescence were read every 10 min, with continuous shaking at 37°C, for 2 hours using a plate reader (Tecan). For all assays, values were reported as OD_{600}-normalized luminescence.

Phage-assisted noncontinuous evolution (PANCE)

PANCE was performed as described previously (*47*). In brief, S2060 host cells transformed with selection plasmids were made chemically competent and transformed with mutagenesis plasmid (MP6) (*48*), then plated on 2×YT agar containing 100 mM glucose and appropriate antibiotics. 8 to 12 colonies were picked into individual wells of a 96-well deep well plate (Axygen) containing 1 ml of DRM and appropriate antibiotics. Colonies were resuspended and serially diluted 10-fold, seven times into DRM. The plate was grown at 37°C with shaking at 230 RPM overnight for 16 to 18 h. Wells containing dilutions with OD_{600} ~0.3 to 0.4 were combined, then treated with 10 mM arabinose to induce mutagenesis. This mixture was distributed into 1-ml cultures in a 96-well deep well plate (Axygen). The cultures were then infected with SP at the indicated dilution (aiming for ~1E5 input phage). Infected cultures were grown overnight for 16 to 18 hours at 37°C and

harvested the next day by centrifugation for 10 min at 3400 *g*. 100 µl of the SP-containing supernatant was transferred to a 96-well PCR plate (Thermo Fisher Scientific), sealed with foil and stored at 4°C. SP were then used to infect the next passage, and the process was repeated for the duration of the selection. Phage titers were determined by qPCR as described previously (*47*) or by the plaque assay described above. If titers were low (<1E4 pfu/ml), a passage of drift was performed (*47*). For drift passages, S2208 encoding MP6 were used instead of selection strains. In drift passages, SP were only allowed to propagate for 6 to 8 hours instead of overnight to minimize the likelihood of recombination of *gIII* into the SP genome. Following completion of a PANCE campaign (upon a noticeable change in phage propagation on the selection strain), SP were plaqued using S2208 cells or the selection strain. The evolved genes of interest from individual plaques were then amplified by PCR, as described previously (*47*), and submitted for Sanger (Quintara Biosciences) or Nanopore (Plasmidsaurus) sequencing to generate inputs for Mutato analyses (https://hub.docker.com/r/araguram/mutato).

Phage-assisted continuous evolution (PACE)

PACE was performed as previously described (*47*). Briefly, host cells containing the mutagenesis plasmid were prepared as described for PANCE above. 12 colonies were picked into individual wells of a 96-well deep well plate (Axygen) containing 1 ml of DRM and appropriate antibiotics. Colonies were resuspended and serially diluted 10-fold, seven times into DRM. The plate was grown at 37°C with shaking at 230 RPM overnight for 16 to 18 hours. Wells containing dilutions with OD_{600} ~0.3 to 0.4 were combined and used to inoculate a chemostat containing 100 l of DRM. The chemostat was grown to OD_{600} ~0.4 to 0.8 and then continuously diluted with fresh DRM at a rate of 1 to 1.5 chemostat volumes per hour to keep cell density constant. The chemostat was maintained at a volume of 80 to 100 ml.

Before SP infection, lagoons were filled with 15 ml of culture from the chemostat and pre-induced with 10 mM arabinose for at least 1 hour. Lagoons were infected with SP at a high starting titer (typically ~10^8 pfu/ml). To increase stringency, the lagoon dilution rates were increased over time as indicated. During the evolution, samples (~500 µl) of the lagoon were collected from the lagoon waste lines at the indicated times. Samples were centrifuged at 4000 *g* for 5 min, and the SP-containing supernatant was stored at 4°C. Titers of SP samples were determined by plaque assays.

Following completion of a PACE campaign (after the lagoon dilution rate was 3 vol/hour for >24 hours), final SP samples were plaqued on the S2060 strain to determine whether *gIII*-recombinant SP had formed during selection [which "cheat" the selection by enabling activity-independent propagation (*47*)]. If cheating was observed (plaques on S2060 strain), individual plaques were amplified overnight in 1 ml DRM at 37°C with shaking at 230 RPM, and then the culture was centrifuged at 4000 *g* for 5 min. The cell pellet, containing SP-infected *E. coli*, was miniprepped to isolate the double-stranded DNA replicative form of the SP (*109*). This isolated SP DNA was then sent for Nanopore sequencing (Plasmidsaurus) to determine the sequence of the *gIII*-recombinant SP, allowing inspection of the mechanism of *gIII* acquisition. If cheating was not detected (no plaques on S2060 strain), final SP samples were plaqued using S2208 cells or the selection strain. The evolved genes of interest from individual plaques were then amplified by PCR, as described previously (*47*), and submitted for Sanger (Quintara Biosciences) or Nanopore (Plasmidsaurus) sequencing to generate inputs for Mutato analyses (https://hub.docker.com/r/araguram/mutato).

General mammalian cell culture conditions

HEK293T (ATCC CRL-3216), HeLa (ATCC CCL-2), and HuH7 (a gift from Erik Sontheimer's group, originated from ATCC) cells were cultured and passaged in Dulbecco's modified Eagle's medium (DMEM) plus GlutaMAX (Thermo Fisher Scientific) supplemented with 10% (v/v) fetal bovine serum (Gibco, qualified). K562 (ATCC CCL-243) cells were cultured and passaged in RPMI medium 1640 plus GlutaMAX (Thermo Fisher Scientific) supplemented with 10% (v/v) fetal bovine serum (Gibco, qualified). All cell types were incubated, maintained, and cultured at 37°C with 5% CO_2. Cell lines were authenticated by their respective suppliers and were negative for mycoplasma by testing with MycoAlert (Lonza Biologics).

Transfection protocol for genome editing in HEK293T cells and genomic DNA preparation

HEK293T cells were seeded on 48-well poly-D-lysine coated plates (Corning) at a density of 40,000 to 45,000 cells per well. HEK293T cells used for seeding were kept under 15 passages since initial thaw from the ATCC vial to ensure high transfectability. 16 to 24 hours after seeding, cells were transfected at 60 to 80% confluency with 1.5 µl Lipofectamine 2000 (Thermo Fisher Scientific).

Initial CAST transfections (figs. S7, S8, and S20A) used the following stoichiometry of components: 50 ng pCas6, 50 ng pCas7, 50 ng pCas8, 50 ng pTniQ, 150 ng pTnsAB, 150 ng pTnsC, and 300 ng pDonor-crRNA. For conditions targeting a plasmid substrate (fig. S20A), 2 ng of pTarget was added. For these initial transfections, cells were cultured for a total of 3 days following transfection.

Following optimization of *Pse*CAST transfection conditions (*44*), a new stoichiometry of components was implemented for all CAST transfections, unless otherwise stated: 50 ng pCas6, 50 ng pCas7, 50 ng pCas8, 50 ng pTniQ, 100 ng pTnsAB, 25 ng pTnsC, 300 ng pDonor-crRNA, and 20 ng pPuroR. ClpX was only delivered if specified, in which case 20 ng pClpX was added. For conditions targeting a plasmid substrate (figs. S17C and S24), 2 ng of pTarget was added. eePASSIGE experiments were performed as described previously (*37*), scaled for a 48-well transfection: 250 ng PEmax plasmid, 37.5 ng of each pegRNA plasmid, 250 ng Bxb1 plasmid, and 375 ng donor plasmid. At 24 hours posttransfection, media was exchanged for fresh DMEM + 10% FBS containing 1 µg/ml puromycin (Thermo Fisher Scientific) to select for transfected cells. Unless otherwise stated, cells were cultured for 3 days following media change, for a total of 4 days incubation posttransfection.

For all experiments, at time of harvest the media was removed, the cells were washed with 1× PBS solution (Thermo Fisher Scientific), and genomic DNA was extracted via the addition of 100 µl of freshly prepared lysis buffer [10 mM Tris-HCl, pH 8.0; 0.05% SDS; 25 ug/ml proteinase K (Thermo Fisher Scientific)] directly into each well of the tissue culture plate. The genomic DNA mixture was incubated at 37°C for >1 hour, followed by an 80°C enzyme inactivation step for 30 min. This lysed mixture containing genomic DNA was stored at −20°C until use.

HTS of genomic DNA samples

HTS was used to quantify integration efficiencies as previously described (*44*), also outlined in fig. S5A, with minor modification. Following genomic DNA isolation, 1 µl of the genomic DNA extract was used as input for the first of two PCR reactions. Genomic loci were amplified in PCR1 using Q5 Hot Start High-Fidelity 2× Master Mix (New England BioLabs). PCR1 primers are listed in table S12. PCR1 was performed as follows: 98°C for 3 min; 25 cycles of 98°C for 15 s, 65°C for 20 s, and 72°C for 30 s; 72°C for 2 min. PCR1 products were confirmed on a 1.5% agarose gel. 1 µl of PCR1 was used as an input for PCR2 to append Illumina barcodes. PCR2 was conducted for 10 cycles of amplification using Q5 Hot Start High-Fidelity 2× Master Mix (New England BioLabs). Following PCR2, samples were pooled and gel purified in a 1.5% agarose gel using a Qiaquick Gel Extraction Kit (Qiagen). For conditions using a unique molecular identifier (UMI) (fig. S32D), a PCR0 step was performed prior to PCR1. PCR0 was performed in a 20-µl reaction using 1 µl genomic DNA extract, 0.1 µM of a 15-nt UMI-containing forward primer (table S12), and Phusion U Hot Start II DNA polymerase (Thermo Fisher Scientific) using the following conditions: 98°C for 3 min; 10 cycles of 98°C for 15 s, 61°C

for 20 s, and 72°C for 1 min. PCR0 products were bead purified using 1.6× AMPure beads (Beckman Coulter) and eluted in 10 µl water. 2 µl of eluted PCR0 product was then used as template for PCR 1, performed as described above. Library concentration was quantified using the Qubit High-Sensitivity Assay Kit (Thermo Fisher Scientific). Samples were sequenced on an Illumina MiSeq instrument (paired-end read, read 1: 200 to 300 cycles, read 2: 0 cycles) using an Illumina MiSeq 300 v2 Kit (Illumina).

Sequencing reads were demultiplexed using MiSeq Reporter (Illumina). A custom Python script was used for quantification of CAST integration efficiencies, which aligned amplicons to either the unedited sequence or integrated sequence. Integration efficiency was calculated as: percentage of (number of integrated reads)/(number of integrated reads + number of unedited reads). This analysis pipeline was also used to determine the distribution of T-RL and T-LR insertion sites downstream of the target site (Fig. 4D and fig. S6, B and C). For detection of indels (Fig. 4E), amplicons were aligned to reference sequences using CRISPResso2 (*110*). For detection of substitutions in fig. S23, amplicons were analyzed using a custom Python script. Sequencing reads for eePASSIGE experiments (fig. S32D) were deduplicated using AmpUMI (*111*) and then analyzed using CRISPResso2 (*110*), as previously described (*37*).

ddPCR quantification of integration efficiency

ddPCR quantification of integration efficiencies was performed as described previously (*44*), also outlined in fig. S4B, with modification. Primer pairs spanned the genome-transposon junction, with probes designed to be specific to the most frequent T-RL integration site, which was determined via HTS (the same was done for T-LR detection for Fig. 4F). Primer pairs and probes are listed in table S12. Because probes could partially hybridize to other integration sites (fig. S6A), the set of integration products that could be detected by each primer pair/probe was determined using mock-integrated standards synthesized as eBlocks (Integrated DNA Technologies) (fig. S6, D and E). For quantification of integration frequencies in genomic DNA samples, 1 µl of crude genomic DNA extract was added to a 25-µl (final volume) reaction mixture containing a final concentration of 1× ddPCR Supermix for Probes (no dUTP) (BioRad), 900 nM of each reference primer, 900 nM of each target primer, 250 nM reference probe (HEX-labeled), 250 nM target probe (FAM-labeled), and 0.2 U/µl HindIII (New England BioLabs). For all assays, the reference primer pair and probe targeted *ACTB* (BioRad, unique assay ID: dHsaCNS141996500), except for *ACTB*-targeting CAST conditions, which used *GAPDH* (BioRad, unique assay ID: dHsaCNS794216737). Droplet generation, PCR, and droplet reading steps were performed using the BioRad QX ONE platform. PCR was performed as follows: 95°C for 10 min; 40 cycles of 94°C for 30 s and 58°C for 2 min; and 98°C for 10 min. Data were analyzed using the BioRad QX ONE software 1.3, Standard Edition, according to the manufacturer's instructions. Integration efficiency was calculated as: percentage of (concentration of integrated molecules)/(concentration of reference molecules). All reported integration efficiencies, unless otherwise noted, were determined using primer pair/probes specific to T-RL integration, which comprised >95% of total integration events for evoCAST (Fig. 4F). As shown in fig. S6F, because T-RL probes often did not detect all possible T-RL integration sites, reported efficiencies are likely underestimates of true integration efficiencies. For eePASSIGE experiments, previously reported (*37*) primer pairs were used (also in table S12), with ddPCR reactions performed and analyzed as described above.

Cell viability assay to assess ClpX toxicity

HEK293T cells were seeded on 96-well poly-D-lysine coated plates (Corning) at a density of ~5000 cells per well. 18 to 24 hours after seeding, cells were transfected at 60 to 80% confluency with 150 ng of plasmid expressing mCherry, ClpX, or ClpX with catalytic inactivating mutations in 1 µl Lipofectamine 2000 (Thermo Fisher Scientific). Each day, starting on the day of transfection (D0), cell viability was measured with the CellTiter-Glo2.0 assay (Promega) according to the manufacturers protocol. Luminescence was measured in 96-well flat-bottomed polystyrene microplates (Corning) using a M1000 Pro microplate reader (Tecan) with a 1-s integration time.

Generation of Δ*clpX E. coli* strain for PACE

Lambda red recombineering was performed as described previously (*57*), with modification, to generate a S1021-derivative (*48*) *E. coli* strain that lacked endogenous *clpX*. In brief, S1021 (*48*) transformed with pKD119 (*112*) were grown to OD_{600} ~0.6 at 30°C with shaking at 230 RPM in SOC media (New England BioLabs) with tetracycline and 2 mM arabinose. Mid-log cells were made electrocompetent via washing with ice-cold 10% glycerol and then electroporated with double-stranded donor DNA containing FRT-KanR-FRT (*112*) and homology arms targeting the flanking genomic regions of *clpX*. Electroporated cells were allowed to recover in 1 ml SOC (New England BioLabs) at 30°C overnight with shaking at 230 RPM. This overnight culture was plated on 2×YT agar with kanamycin and incubated at 30°C overnight. An individual recombinant colony, confirmed via PCR and Sanger sequencing (Quintara Biosciences), was grown overnight in 2×YT with kanamycin at 37°C with shaking at 230 RPM to cure the temperature-sensitive pKD119 plasmid. These cells were then made chemically competent and transformed with pBAD-Flp (*113*) (a gift from Lydia Freddolino, Addgene plasmid # 122969). Transformed cells were plated on 2×YT agar with chloramphenicol and 10 mM arabinose, and incubated at 30°C overnight to allow the KanR cassette to recombine out of the genome. An individual colony that successfully recombined out the KanR cassette was identified via PCR and was then grown overnight in 2×YT at 37°C with shaking at 230 RPM to cure the temperature-sensitive pBAD-Flp. This overnight culture was plated on 2×YT agar containing streptomycin (selecting for the strain) and incubated overnight at 37°C. Individual Δ*clpX* colonies were isolated and confirmed to be sensitive to tetracycline and chloramphenicol (confirming pKD119 and pBAD-Flp were cured, respectively). Finally, to conjugate the F plasmid into the newly generated Δ*clpX* strain, a mid-log (OD_{600} ~0.6) culture of F′ donor *E. coli* was diluted 1:1000 into a mid-log (OD_{600} ~0.6) culture of the newly generated F-Δ*clpX* strain and incubated for 1.5 hours at 37°C. This culture was then plated on 2×YT agar with streptomycin (selecting for strain) and tetracycline (selecting for F plasmid), and incubated overnight. An individual F′ Δ*clpX* colony was grown overnight at 37°C in 2×YT with tetracycline and streptomycin, and the overnight culture was used to generate a glycerol stock. Primers used for strain generation are in table S12.

HEK293T fluorescent reporter assay for transposon-end binding and flow cytometry analysis

Transposon-end binding transcriptional activation assays were performed as previously described (*44*). In brief, HEK293T cells were seeded on 48-well poly-D-lysine coated plates (Corning) at a density of 40,000 to 45,000 cells per well. 16 to 24 hours after seeding, cells were transfected at 60 to 80% confluency using 1.5 µl Lipofectamine 2000 (Thermo Fisher Scientific). TnsB variants, fused at the C terminus to VP64, were individually co-transfected with a GFP transfection marker and a reporter plasmid containing a *Pse*CAST transposon end adjacent to a minimal CMV promoter and a tdTomato marker at a ratio of 200 ng:20 ng:60 ng (TnsB:GFP:reporter). 48 to 72 hours post-transfection, cells were analyzed via flow cytometry on a Novocyte Penteon. GFP positive cells were analyzed for tdTomato fluorescence. The bulk mean fluorescence intensity (MFI) was calculated for each transfection and normalized to a transfection in which no TnsB-VP64 transcriptional activator was added.

Western immunoblotting

Western immunoblotting assays were performed as previously described (*44*). In brief, HEK293T cells were seeded on 48-well poly-D-lysine coated

plates (Corning) at a density of 40,000 to 45,000 cells per well. 16 to 24 hours after seeding, cells were transfected at 60 to 80% confluency using 1.5 µl Lipofectamine 2000 (Thermo Fisher Scientific). TnsAB variants were cloned with an internal 3xFLAG-bipartite-NLS fusion, and 200 ng of each variant was individually transfected into individual wells. 48 to 72 hours posttransfection, cells were lysed in lysis buffer [150 mM NaCl, 0.1% Triton X-100, 50 mM Tris-HCl (pH 8.0), cOmplete EDTA-free protease inhibitor (Roche)]. Proteins were resolved by SDS-PAGE and transferred to a PVDF membrane (Thermo Fisher Scientific). The membrane was then washed with TBS-T [50 mM Tris-Cl (pH 7.5), 150 mM NaCl, 0.1% Tween 20] and blocked with blocking buffer (TBS-T with 5% w/v BSA). Membranes were stained with either anti-FLAG M2 antibody (Sigma F3165, diluted 1:10000) or β-Actin antibody (Cell Signaling Technology #3700, diluted 1:10000) overnight at 4°C under gentle rotation. Membranes were then stained with HRP-conjugated secondary antibodies (ab97240, ab97250; diluted 1:10000) at room temperature for 1 hour. Membranes were washed and developed with SuperSignal West Dura (Thermo Fisher Scientific). Band intensities were quantified using Image Lab (BioRad), and the solubility of TnsAB variants was determined by dividing FLAG intensities by β-Actin intensities. Solubilities were normalized to that of wild-type *Pse*TnsAB.

Long-read sequencing of integration products

Detection of cointegrates (fig. S24) followed a previously described protocol with minor modification (*79*). HEK293T cells were transfected as described above for CAST integration assays. Approximately 96 hours posttransfection, cells were lysed and DNA was harvested as previously described. Two separate PCRs were performed with equivalent volumes of input lysate, with primer sequences listed in table S12. The first PCR reaction contained a primer that annealed to the pTarget upstream of the target sequence ("P1" in fig. S24A) and a primer that annealed to the left transposon end ("P2" in fig. S24A). The second PCR contained the same forward primer (P1) but contained a reverse primer that annealed to the pDonor backbone downstream of the transposon ("P3", fig. S24A), such that only cointegrates should be amplified. PCRs were then pooled and purified via 1× magnetic bead cleanup (Omega). Purified samples were then prepped for Nanopore sequencing using the Native Barcoding Kit (Oxford Nanopore, SQK-NBD114.24) and loaded onto a R10.4.1 flow cell, sequencing for 18 to 24 hours. Reads were analyzed using BBDuk from the BBTools suite (v.38.00; https://sourceforge.net/projects/bbmap). Reads with a minimum quality score of 8 were filtered to contain the upstream target region, the right transposon end, and the left transposon end. Filtered that contained the pDonor backbone sequence were considered a cointegrate sequence, while filtered reads that did not contain this sequence were considered a simple insertion. pDonor-backbone containing reads were also aligned to the expected sequence of a cointegrate and manually inspected to ensure accuracy (fig. S24A). The frequency of cointegrates was calculated as percentage of (number of cointegrate reads)/(number of simple insertion reads – number of cointegrate reads). The denominator corrects for the double-counting of cointegrate products as simple insertion reads, since the cointegrate product amplifies with both primer pairs. Analyses of transfections with defined ratios of plasmids containing mock simple insertion and cointegrate products generated a standard curve (fig S24B) that was used to calculate cointegrate product frequencies for experimental conditions.

For long-read sequencing comparing evoCAST and eePASSIGE integration products (fig. S32, G and H), PCR reactions were first purified via 0.9× magnetic bead cleanup (Omega). Purified samples were then prepped for Nanopore sequencing using the Native Barcoding Kit (Oxford Nanopore SQK-NBD114.96). Samples were loaded onto a R10.4.1 flow cell, sequencing for 18 to 24 hours using the super-accurate base calling settings. Samples were processed using porechop with default settings and a minimum split read length of 500 bp. Reads containing the primer binding sites used for PCR amplification (with a Hamming distance <2) were extracted and aligned to a reference locus of either a mock CAST integrant (T-RL, 49-bp insertion site) or a mock PASSIGE integrant using minimap2 (*114*) and the parameters "-ax map-ont–sam-hit-only–secondary=no". The resulting bam files were analyzed for insertions and deletions (indels) via the CIGAR string of each read alignment, and the proportion of indels was calculated relative to the total number of reads at a position. The background sequencing error frequency at each position was calculated using PCR amplicons generated from synthetic mock-integrated fragments (IDT eBlocks). This error frequency was subtracted from the calculated indel frequencies. In the region of the integration product that was covered by both PCR amplicons (fig. S32G), the average indel proportion of both amplicons was reported.

UDiTaS sample preparation, sequencing, and computational analysis

HEK293T cells were transfected as described above for CAST integration assays, except the pPuroR was omitted, and the transposon in the pDonor was modified to contain a promoter-driven puromycin cassette and an N10 UMI immediately flanking the transposon right end (table S11). Following transfection, cells were passaged under puromycin selection for 7 days to enrich the integrated cell population and dilute unintegrated pDonor molecules. Cells were then lysed as described for CAST integration assays, and the genomic DNA (gDNA) was purified via bead cleanup (Omega) and quantified using the Qubit High-Sensitivity Assay Kit (Thermo Fisher Scientific).

TnY was purified as previously described (*115*), preloaded with full-length Nextera Read 2/Indexed oligos, and diluted to the appropriate working concentration such that 100 ng of gDNA would be tagmented into ~2-kb fragments. 100 ng of gDNA were tagmented as previously described (*81*) with modification: following tagmentation, reactions were incubated with 0.4 U of Proteinase K (NEB) for 10 min at 55°C to ensure release of the transposase from the gDNA. Reactions were then column purified with a DNA Clean & Concentrator Kit (Zymo), and eluted in 25 µl. An initial PCR1 was performed with a forward primer that anneals to the transposon cargo upstream of the N10 UMI, and a reverse primer that anneals to the P7 adapter sequence installed via tagmentation (table S12). PCR1 was performed using KAPA HiFi Hotstart (Roche) as follows: 98°C for 5 min; 15 cycles of 98°C for 20 s, 55°C for 30 s, and 72 for 1 min; and 72°C for 5 min. PCR1s were purified using 0.9× Omega Mag-Bind TotalPure magnetic beads (Omega Bio-Tek) and eluted into 50-µl nuclease-free water. 2-µl eluted DNA was used as input for PCR2, which appended Illumina sequencing adapters. After 15 cycles of PCR2 (same conditions as PCR 1), the reaction was resolved on a gel, and a smear corresponding to a size range of 350 to 800 bp was extracted. Samples were sequenced on an Element Biosciences AVITI instrument (paired-end read, read 1: 150 cycles, read 2: 150 cycles) using a Cloudbreak Freestyle Kit (Element Biosciences).

Reads were processed using a custom Python script (*116*). In brief, reads were first trimmed and quality filtered using cutadapt (*117*) (v4.2, -a CTGTCTCTTATACACATCT -A CTGTCTCTTATACACATCT–minimum-length 15 -q 20). After adapter trimming, reads were then filtered to contain the right transposon end using BBDuk from the BBTools suite (v.38.00; https://sourceforge.net/projects/bbmap). Reads aligning to transfected plasmids with a Hamming distance <3 were discarded. An overwhelming fraction of total reads (often >99%) aligned to unintegrated pDonor molecules, limiting coverage of genomic integration events. UMIs were then extracted using umitools (*118*) (v1.1.4, extract–bc-pattern=NNNNNNNNNN). Flank sequences that passed filtering were then mapped to the GRCh38 reference genome using Bowtie2 (*119*) (v2.4.2,–very-sensitive –no-mixed–no-discordant). Alignments were UMI processed using umitools (dedup) (*118*). Lastly, UMI-processed alignments were manually inspected, and insertion events were defined as meeting the following criteria: >1 mapped read per UMI, ≤3 mismatches in the genomic alignment, paired reads mapped within 1200 bp of each other, and a primary read alignment. Insertion events were considered on-target if they occurred <100 bp downstream of the target site.

DNA from *E. coli* PACE host cells was prepared for UDiTaS by resuspending pelleted host cells in water, incubating at 95°C for 10 min, and purifying DNA via bead cleanup (Omega). DNA was quantified using the Qubit High-Sensitivity Assay Kit (Thermo Fisher Scientific), and 100 ng of DNA was tagmented and analyzed as described above, except alignment was to a custom reference sequence containing the *E. coli* DH10B genome (*120*) [from which the S2060 PACE strain is derived (*47*)] and all PACE selection circuit plasmid sequences.

Analysis of ATAC-seq data

ATAC-seq data for HEK293 cells were obtained from GSE108513 (*121*) in BigWig (.bw) format and converted to BedGraph format using the UCSC Genome Browser utility BigWigToBedGraph. For characterization of off-target events (fig. S25D), the 200-bp sequences surrounding each of the mapped off-target integration sites (previously aligned to the GRCh38 genome) were re-aligned to the GRCh37 genome (used in the published ATAC-seq dataset) using Bowtie2 with "–end-to-end" settings, with 55 of 56 off-targets successfully aligning. Off-target locations and the 1-kb surrounding sequences were then extracted using samtools, and the average ATAC-seq score for each 1-kb region was fetched using the bedtools map function (*122*). The average ATAC-seq score of all off-target regions was then calculated. To perform a bootstrap analysis of randomly sampling 1-kb regions of the GRCh37 genome, the bedtools random function was used to generate 20,000 unique sets of 55 1-kb regions. The ATAC-seq scores for the 55 regions within each set were fetched using bedtools map (*122*), and the averages of each set were used to generate the histogram shown in fig. S25D.

To assess the relationship between ATAC-seq signal and on-target evoCAST integration efficiencies (fig. S27), the average ATAC-seq score for each 1-kb region surrounding the target site was fetched using the bedtools map function (*122*). Pearson correlation analysis was performed using Prism 10 (GraphPad) to assess the relationship between on-target integration efficiency and the corresponding mean ATAC-seq score.

HEK293T fluorescent reporter assay for off-target integration and flow cytometry analysis

A reporter assay for off-target integration in HEK293T cells was performed as described previously (*37*), with minor modifications. Briefly, HEK293T cells were seeded on 48-well poly-D-lysine coated plates (Corning) at a density of 40,000 to 45,000 cells per well. 16 to 24 hours after seeding, cells were transfected at 60 to 80% confluency using 1.5 µl Lipofectamine 2000 (Thermo Fisher Scientific). TnsABC conditions contained 300 ng pDonor, 100 ng pTnsAB, and 25 ng pTnsC. Bxb1 conditions contained 450 ng pDonor and 300 ng pBxb1 [plasmid amounts were kept consistent with previously optimized amounts for efficient CAST (*44*) and eePASSIGE (*37*) activity]. For all conditions, the donor contained mCherry under a cytomegalovirus (CMV) promoter. Cells were passaged for 2 weeks posttransfection to dilute the pDonor. Cells were then trypsinized, resuspended in 1× PBS solution, and analyzed via flow cytometry using the CytoFLEX LX Flow Cytometer (Beckman Coulter) at the Broad Institute flow cytometry core.

Quantification of transgene expression via RT-ddPCR

To assess transgene expression, mRNA was isolated from HEK293T cells or HuH7 cells 4 days posttransfection using the RNeasy Plus kit (Qiagen). 400 to 800 ng of isolated RNA were treated with RQ1 RNase-free DNase (Promega) for 1 hour at 37°C in a 10-µl reaction, and then combined with 1 µl RQ1 DNase Stop Solution (Promega) and incubated at 60°C for 10 min. 9 µl of DNase-treated RNA was used as input for a 20-µl reverse transcription reaction containing the SuperScript IV Vilo Master Mix (Thermo Fisher Scientific), performed according to manufacturer's protocols. 1 µl of the reverse transcription reaction was used as input for a 25-µl (final volume) ddPCR reaction containing a final concentration of 1× ddPCR Supermix for Probes (no dUTP) (BioRad), 900 nM of each reference primer, 900 nM of each target primer, 250 nM reference probe (HEX-labeled), 250 nM target probe (FAM-labeled), and 0.2 U/µl HindIII (New England BioLabs). For *MECP2* quantification in HEK293T cells, the target primer pair/probe was designed to be specific to the exon 1–exon 2 junction, where exon 2 of the integrated transgene was recoded such that the target primer pair/probe did not detect endogenous *MECP2* expression. For *F9* quantification in HuH7 cells, the target primer pair/probe was designed to be specific to the *ALB* exon 1–*F9* exon 2 junction. Primer pairs and probes are listed in table S12. The reference primer pair and probe targeted *TBP* (BioRad, unique assay ID: dHsaCPE5058363). ddPCR and analysis were performed as described for the quantification of integration efficiencies. Transgene expression was reported as the concentration of target transcripts divided by the concentration of *TBP* transcripts.

Generation of a clonal evoCAST-edited HEK293T cell line

HEK293T cells were transfected according to the above protocol for CASTs in HEK293T cells, except pPuroR was omitted. To enable selection for edited cells, the transposon contained a splice acceptor and puromycin resistance gene (table S11), such that integration into the transcriptionally active *AAVS1* locus would enable puromycin resistance. Cells were placed under selection with 1 µg/ml puromycin (Thermo Fisher Scientific) 4 days posttransfection and passaged for >1 month (this long duration was chosen to demonstrate the durability of CAST-edited cells in a bulk-transfected population). Cells were then single-cell sorted into poly-D-lysine coated 96-well plates (Corning) using a MA900 Cell Sorter (Sony) with the single cell 3-drop setting. Sorted cells were monitored after sorting, and wells with single colonies were marked for further analysis. After the cells had expanded for ~10 days, marked wells were split into two separate poly-D-lysine coated 96-well plates (Corning). After additional expansion for 3 to 5 days, one plate of the expanded cells was harvested for analysis of integration efficiency by ddPCR. Clonal cell lines with detectable integration events via ddPCR were further expanded for cell line generation.

Transfection of HeLa and HuH7 cells

For HeLa cell transfections, cells were seeded on 48-well poly-D-lysine coated plates (Corning) at a density of 30,000 cells per well. Between 16 to 24 hours after seeding, cells were transfected at 60 to 80% confluency with 1 µl Lipofectamine 3000 (Thermo Fisher Scientific). For HuH7 cell transfections, cells were seeded on 48-well poly-D-lysine coated plates (Corning) at a density of 40,000 cells per well. Between 16 to 24 hours after seeding, cells were transfected at 60 to 80% confluency with 1.5 µl Lipofectamine 2000 (Thermo Fisher Scientific). For both HeLa and Huh7 cells, transfections used 50 ng pCas6, 50 ng pCas7, 50 ng pCas8, 50 ng pTniQ, 150 ng pTnsAB, 25 ng pTnsC, 300 ng pDonor-crRNA, and 20 ng pPuroR. For HeLa transfections with ClpX, 20 ng pClpX was used. At 24 hours posttransfection, media was exchanged for fresh DMEM + 10% FBS containing 1 ug/ml puromycin (Thermo Fisher Scientific) to select for transfected cells. Cells were cultured for 3 days following media change, for a total of 4 days incubation posttransfection. Genomic DNA isolation was performed as described above for HEK293T transfections.

Nucleofection of K562 cells

For K562 nucleofections, 3 µg of CAST components [same ratio of components as used in HEK293T cell experiments: 215 ng pCas6, 215 ng pCas7, 215 ng pCas8, 215 ng pTniQ, 647 ng pTnsAB, 108 ng pTnsC, 1.29 µg pDonor-crRNA, 86 ng pPuroR, and 86 ng pClpX (if used)] were nucleofected in a final volume of 20 µl in a 16-well nucleocuvette strip (Lonza). Cells were nucleofected using the SE Cell Line 4D-Nucleofector X Kit (Lonza), with 500,000 cells per sample (program FF-120), according to the manufacturer's protocol. At 24 hours posttransfection, 1 ug/ml puromycin (Thermo Fisher Scientific) was added to wells to select for transfected cells. Cells were cultured for 3 days following media change, for a total of 4 days incubation posttransfection. Genomic DNA isolation was performed as described above for HEK293T transfections.

Primary human fibroblast cell culture conditions and electroporation

Following the Declaration of the Helsinki Principles, primary dermal fibroblast cells were obtained from a recessive dystrophic epidermolysis bullosa (RDEB) patient via a 3-mm punch biopsy. Tissue was minced and plated in a 6-well dish. Adherent cells were expanded and cultured in MEM-alpha complete media containing 20% (v/v) fetal bovine serum (Atlas Biologicals, F-0500-D), 1× Glutamax, 1× penicillin/streptomycin (Thermo Fisher 15-140-122), 1× Antioxidant Supplement (Sigma Aldrich, A1345), 1× nonessential amino acids (Thermo Scientific, 11140050), 10 ng/ml epidermal growth factor (Sigma Aldrich, E4127), and 0.5 ng/ml fibroblast growth factor (Sigma Aldrich F3133). Fibroblasts were passaged when at 100% confluency. For evoCAST experiments, cells were cultured in a T-150 flask (Fisher Scientific 1012634) and harvested by trypsinization using Trypsin-EDTA (0.25%; Thermo Scientific 25200114). Cells were removed using complete media, and then washed and resuspended in Neon Buffer R.

Per condition, two electroporations of 100,000 cells each were performed using the Neon NXT instrument (Thermo Fisher) with the following protocol: 1700 V, 20 ms, 1 pulse, using buffers R and E10 with 10-µl tips. Each electroporation (per 100,000 cells) contained: 200 ng pCas6, 200 ng pCas7, 200 ng pCas8, 200 ng pTniQ, 800 ng pTnsAB, 200 ng pTnsC, 582 ng linearized donor-crRNA, and 80 ng pClpX (if added). Donor-crRNA was delivered as linear DNA to mitigate total DNA mass and related DNA toxicity. Both electroporation reactions were combined into one well of a 24-well plate (thus constituting a single replicate) containing MEM-alpha complete media as above, but without penicillin and streptomycin. Complete, conditioned media from parental fibroblasts was added ~4 hours postelectroporation and then replenished daily until harvest. Cells were harvested 5 days postelectroporation. Media was first removed, then cells were trypsinized (Trypsin-EDTA (0.25%); Thermo Scientific 25200114) and harvested in the MEM-alpha media. Resuspended cells were pelleted by centrifugation, and genomic DNA was isolated using the Monarch Genomic DNA Purification Kit (New England BioLabs, T3010S). Maintaining high cell density postelectroporation was critical to preserving cell viability and observing higher editing, as the above electroporation protocol using a 12-well plate, rather than a 24-well plate, led to substantially fewer viable cells and an associated lower editing (<1%).

To assess viability, cells were imaged 48 hours postelectroporation with a Leica DMiI microscope (Thomas Scientific) using the 5× magnification objective.

REFERENCES AND NOTES

1. M. Jinek *et al.*, A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012). doi: 10.1126/science.1225829; pmid: 22745249
2. L. Cong *et al.*, Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013). doi: 10.1126/science.1231143; pmid: 23287718
3. P. Mali *et al.*, RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013). doi: 10.1126/science.1232033; pmid: 23287722
4. A. V. Anzalone, L. W. Koblan, D. R. Liu, Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.* **38**, 824–844 (2020). doi: 10.1038/s41587-020-0561-9; pmid: 32572269
5. A. C. Komor, Y. B. Kim, M. S. Packer, J. A. Zuris, D. R. Liu, Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016). doi: 10.1038/nature17946; pmid: 27096365
6. N. M. Gaudelli *et al.*, Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017). doi: 10.1038/nature24644; pmid: 29160308
7. B. Y. Mok *et al.*, A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. *Nature* **583**, 631–637 (2020). doi: 10.1038/s41586-020-2477-4; pmid: 32641830
8. T. P. Huang, G. A. Newby, D. R. Liu, Precision genome editing using cytosine and adenine base editors in mammalian cells. *Nat. Protoc.* **16**, 1089–1128 (2021). doi: 10.1038/s41596-020-00450-9; pmid: 33462442
9. A. V. Anzalone *et al.*, Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019). doi: 10.1038/s41586-019-1711-4; pmid: 31634902
10. J. L. Doman *et al.*, Phage-assisted evolution and protein engineering yield compact, efficient prime editors. *Cell* **186**, 3983–4002.e26 (2023). doi: 10.1016/j.cell.2023.07.039; pmid: 37657419
11. J. Yan *et al.*, Improving prime editing with an endogenous small RNA-binding protein. *Nature* **628**, 639–647 (2024). doi: 10.1038/s41586-024-07259-6; pmid: 38570691
12. P. J. Chen, D. R. Liu, Prime editing for precise and highly versatile genome manipulation. *Nat. Rev. Genet.* **24**, 161–177 (2023). doi: 10.1038/s41576-022-00541-1; pmid: 36344749
13. H. Frangoul *et al.*, Exagamglogene Autotemcel for Severe Sickle Cell Disease. *N. Engl. J. Med.* **390**, 1649–1662 (2024). doi: 10.1056/NEJMoa2309676; pmid: 38661449
14. R. G. Lee *et al.*, Efficacy and Safety of an Investigational Single-Course CRISPR Base-Editing Therapy Targeting *PCSK9* in Nonhuman Primate and Mouse Models. *Circulation* **147**, 242–253 (2023). doi: 10.1161/CIRCULATIONAHA.122.062132; pmid: 36314243
15. R. Chiesa *et al.*, Base-Edited CAR7 T Cells for Relapsed T-Cell Acute Lymphoblastic Leukemia. *N. Engl. J. Med.* **389**, 899–910 (2023). doi: 10.1056/NEJMoa2300709; pmid: 37314354
16. J. D. Gillmore *et al.*, CRISPR-Cas9 In Vivo Gene Editing for Transthyretin Amyloidosis. *N. Engl. J. Med.* **385**, 493–502 (2021). doi: 10.1056/NEJMoa2107454; pmid: 34215024
17. US National Institutes of Health, ClinicalTrials.gov (2024); https://clinicaltrials.gov/.
18. Q. Xiao, V. M. Lauschke, The prevalence, genetic complexity and population-specific founder effects of human autosomal recessive disorders. *NPJ Genom. Med.* **6**, 41 (2021). doi: 10.1038/s41525-021-00203-x; pmid: 34078906
19. J. T. Bulcha, Y. Wang, H. Ma, P. W. L. Tai, G. Gao, Viral vector platforms within the gene therapy landscape. *Signal Transduct. Target. Ther.* **6**, 53 (2021). doi: 10.1038/s41392-021-00487-6; pmid: 33558455
20. R. Zufferey *et al.*, Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery. *J. Virol.* **72**, 9873–9880 (1998). doi: 10.1128/JVI.72.12.9873-9880.1998; pmid: 9811723
21. J. Earley, E. Piletska, G. Ronzitti, S. Piletsky, Evading and overcoming AAV neutralization in gene therapy. *Trends Biotechnol.* **41**, 836–845 (2023). doi: 10.1016/j.tibtech.2022.11.006; pmid: 36503641
22. Assessment of adenoviral vector safety and toxicity: Report of the National Institutes of Health Recombinant DNA Advisory Committee. *Hum. Gene Ther.* **13**, 3–13 (2002). doi: 10.1089/10430340152712629; pmid: 11779411
23. M. G. Sacco *et al.*, Lymphoid abnormalities in CD40 ligand transgenic mice suggest the need for tight regulation in gene therapy approaches to hyper immunoglobulin M (IgM) syndrome. *Cancer Gene Ther.* **7**, 1299–1306 (2000). doi: 10.1038/sj.cgt.7700232; pmid: 11059686
24. M. Van Alstyne *et al.*, Gain of toxic function by long-term AAV9-mediated SMN overexpression in the sensorimotor circuit. *Nat. Neurosci.* **24**, 930–940 (2021). doi: 10.1038/s41593-021-00827-3; pmid: 33795885
25. D. Pannell, J. Ellis, Silencing of gene expression: Implications for design of retrovirus vectors. *Rev. Med. Virol.* **11**, 205–217 (2001). doi: 10.1002/rmv.316; pmid: 11479927
26. P. Rouet, F. Smih, M. Jasin, Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol. Cell. Biol.* **14**, 8096–8106 (1994). pmid: 7969147
27. K. Suzuki *et al.*, In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature* **540**, 144–149 (2016). doi: 10.1038/nature20565; pmid: 27851729
28. W. D. Heyer, K. T. Ehmsen, J. Liu, Regulation of homologous recombination in eukaryotes. *Annu. Rev. Genet.* **44**, 113–139 (2010). doi: 10.1146/annurev-genet-051710-150955; pmid: 20690856
29. M. Kosicki, K. Tomberg, A. Bradley, Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* **36**, 765–771 (2018). doi: 10.1038/nbt.4192; pmid: 30010673
30. E. Haapaniemi, S. Botla, J. Persson, B. Schmierer, J. Taipale, CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nat. Med.* **24**, 927–930 (2018). doi: 10.1038/s41591-018-0049-z; pmid: 29892067
31. B. Chaikind, J. L. Bessen, D. B. Thompson, J. H. Hu, D. R. Liu, A programmable Cas9-serine recombinase fusion protein that operates on DNA sequences in mammalian cells. *Nucleic Acids Res.* **44**, 9758–9770 (2016). doi: 10.1093/nar/gkw707; pmid: 27515511
32. B. E. Hew, R. Sato, D. Mauro, I. Stoytchev, J. B. Owens, RNA-guided *piggyBac* transposition in human cells. *Synth. Biol.* **4**, ysz018 (2019). doi: 10.1093/synbio/ysz018; pmid: 31355344
33. S. P. Chen, H. H. Wang, An Engineered Cas-Transposon System for Programmable and Site-Directed DNA Transpositions. *CRISPR J.* **2**, 376–394 (2019). doi: 10.1089/crispr.2019.0030; pmid: 31742433
34. A. Kovač *et al.*, RNA-guided retargeting of *Sleeping Beauty* transposition in human cells. *eLife* **9**, e53868 (2020). doi: 10.7554/eLife.53868; pmid: 32142408
35. A. V. Anzalone *et al.*, Programmable deletion, replacement, integration and inversion of large DNA sequences with twin prime editing. *Nat. Biotechnol.* **40**, 731–740 (2022). doi: 10.1038/s41587-021-01133-w; pmid: 34887556
36. M. T. N. Yarnall *et al.*, Drag-and-drop genome insertion of large sequences without double-strand DNA cleavage using CRISPR-directed integrases. *Nat. Biotechnol.* **41**, 500–512 (2023). doi: 10.1038/s41587-022-01527-4; pmid: 36424489
37. S. Pandey *et al.*, Efficient site-specific integration of large genes in mammalian cells via continuously evolved recombinases and prime editing. *Nat. Biomed. Eng.* **9**, 22–39 (2025). doi: 10.1038/s41551-024-01227-1; pmid: 38858586
38. J. E. Peters, K. S. Makarova, S. Shmakov, E. V. Koonin, Recruitment of CRISPR-Cas systems by Tn7-like transposons. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E7358–E7366 (2017). doi: 10.1073/pnas.1709035114; pmid: 28811374
39. S. E. Klompe, P. L. H. Vo, T. S. Halpin-Healy, S. H. Sternberg, Transposon-encoded CRISPR-Cas systems direct RNA-guided DNA integration. *Nature* **571**, 219–225 (2019). doi: 10.1038/s41586-019-1323-z; pmid: 31189177
40. J. Strecker *et al.*, RNA-guided DNA insertion with CRISPR-associated transposases. *Science* **365**, 48–53 (2019). doi: 10.1126/science.aax9181; pmid: 31171706

41. P. L. H. Vo, C. Acree, M. L. Smith, S. H. Sternberg, Unbiased profiling of CRISPR RNA-guided transposition products by long-read sequencing. *Mob. DNA* **12**, 13 (2021). doi: 10.1186/s13100-021-00242-2; pmid: 34103093
42. P. L. H. Vo *et al.*, CRISPR RNA-guided integrases for high-efficiency, multiplexed bacterial genome engineering. *Nat. Biotechnol.* **39**, 480–489 (2021). doi: 10.1038/s41587-020-00745-y; pmid: 33230293
43. S. E. Klompe *et al.*, Evolutionary and mechanistic diversity of Type I-F CRISPR-associated transposons. *Mol. Cell* **82**, 616–628.e5 (2022). doi: 10.1016/j.molcel.2021.12.021; pmid: 35051352
44. G. D. Lampe *et al.*, Targeted DNA integration in human cells without double-strand breaks using CRISPR-associated transposases. *Nat. Biotechnol.* **42**, 87–98 (2024). doi: 10.1038/s41587-023-01748-1; pmid: 36991112
45. G. Bourque *et al.*, Ten things you should know about transposable elements. *Genome Biol.* **19**, 199 (2018). doi: 10.1186/s13059-018-1577-z; pmid: 30454069
46. K. M. Esvelt, J. C. Carlson, D. R. Liu, A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499–503 (2011). doi: 10.1038/nature09929; pmid: 21478873
47. S. M. Miller, T. Wang, D. R. Liu, Phage-assisted continuous and non-continuous evolution. *Nat. Protoc.* **15**, 4101–4127 (2020). doi: 10.1038/s41596-020-00410-3; pmid: 33199872
48. A. H. Badran, D. R. Liu, Development of potent in vivo mutagenesis plasmids with broad mutational spectra. *Nat. Commun.* **6**, 8425 (2015). doi: 10.1038/ncomms9425; pmid: 26443021
49. B. C. Dickinson, A. M. Leconte, B. Allen, K. M. Esvelt, D. R. Liu, Experimental interrogation of the path dependence and stochasticity of protein evolution using phage-assisted continuous evolution. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 9007–9012 (2013). doi: 10.1073/pnas.1220670110; pmid: 23674678
50. A. M. Leconte *et al.*, A population-based experimental model for protein evolution: Effects of mutation rate and selection stringency on evolutionary outcomes. *Biochemistry* **52**, 1490–1499 (2013). doi: 10.1021/bi3016185; pmid: 23360105
51. J. C. Carlson, A. H. Badran, D. A. Guggiana-Nilo, D. R. Liu, Negative selection and stringency modulation in phage-assisted continuous evolution. *Nat. Chem. Biol.* **10**, 216–222 (2014). doi: 10.1038/nchembio.1453; pmid: 24487694
52. B. C. Dickinson, M. S. Packer, A. H. Badran, D. R. Liu, A system for the continuous directed evolution of proteases rapidly reveals drug-resistance mutations. *Nat. Commun.* **5**, 5352 (2014). doi: 10.1038/ncomms6352; pmid: 25355134
53. M. S. Packer, H. A. Rees, D. R. Liu, Phage-assisted continuous evolution of proteases with altered substrate specificity. *Nat. Commun.* **8**, 956 (2017). doi: 10.1038/s41467-017-01055-9; pmid: 29038472
54. T. R. Blum *et al.*, Phage-assisted evolution of botulinum neurotoxin proteases with reprogrammed specificity. *Science* **371**, 803–810 (2021). doi: 10.1126/science.abf5972; pmid: 33602850
55. A. H. Badran *et al.*, Continuous evolution of Bacillus thuringiensis toxins overcomes insect resistance. *Nature* **533**, 58–63 (2016). doi: 10.1038/nature17938; pmid: 27120167
56. T. Wang, A. H. Badran, T. P. Huang, D. R. Liu, Continuous directed evolution of proteins with improved soluble expression. *Nat. Chem. Biol.* **14**, 972–980 (2018). doi: 10.1038/s41589-018-0121-5; pmid: 30127387
57. M. S. Morrison, T. Wang, A. Raguram, C. Hemez, D. R. Liu, Disulfide-compatible phage-assisted continuous evolution in the periplasmic space. *Nat. Commun.* **12**, 5959 (2021). doi: 10.1038/s41467-021-26279-8; pmid: 34645844
58. A. K. Brödel, R. Rodrigues, A. Jaramillo, M. Isalan, Accelerated evolution of a minimal 63-amino acid dual transcription factor. *Sci. Adv.* **6**, eaba2728 (2020). doi: 10.1126/sciadv.aba2728; pmid: 32577520
59. J. A. M. Mercer *et al.*, Continuous evolution of compact protein degradation tags regulated by selective molecular glues. *Science* **383**, eadk4422 (2024). doi: 10.1126/science.adk4422; pmid: 38484051
60. T. B. Roth, B. M. Woolston, G. Stephanopoulos, D. R. Liu, Phage-Assisted Evolution of Bacillus methanolicus Methanol Dehydrogenase 2. *ACS Synth. Biol.* **8**, 796–806 (2019). doi: 10.1021/acssynbio.8b00481; pmid: 30856338
61. J. H. Hu *et al.*, Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57–63 (2018). doi: 10.1038/nature26155; pmid: 29512652
62. S. M. Miller *et al.*, Continuous evolution of SpCas9 variants compatible with non-G PAMs. *Nat. Biotechnol.* **38**, 471–481 (2020). doi: 10.1038/s41587-020-0412-8; pmid: 32042170
63. B. W. Thuronyi *et al.*, Continuous evolution of base editors with expanded target compatibility and improved activity. *Nat. Biotechnol.* **37**, 1070–1079 (2019). doi: 10.1038/s41587-019-0193-0; pmid: 31332326
64. M. F. Richter *et al.*, Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat. Biotechnol.* **38**, 883–891 (2020). doi: 10.1038/s41587-020-0453-z; pmid: 32433547
65. B. Y. Mok *et al.*, CRISPR-free base editors with enhanced activity and expanded targeting scope in mitochondrial and nuclear DNA. *Nat. Biotechnol.* **40**, 1378–1387 (2022). doi: 10.1038/s41587-022-01256-8; pmid: 35379961
66. M. E. Neugebauer *et al.*, Evolution of an adenine base editor into a small, efficient cytosine base editor with low off-target activity. *Nat. Biotechnol.* **41**, 673–685 (2023). doi: 10.1038/s41587-022-01533-6; pmid: 36357719
67. T. P. Huang *et al.*, High-throughput continuous evolution of compact Cas9 variants targeting single-nucleotide-pyrimidine PAMs. *Nat. Biotechnol.* **41**, 96–107 (2023). doi: 10.1038/s41587-022-01410-2; pmid: 36076084
68. B. P. Hubbard *et al.*, Continuous directed evolution of DNA-binding proteins to improve TALEN specificity. *Nat. Methods* **12**, 939–942 (2015). doi: 10.1038/nmeth.3515; pmid: 26258293
69. L. Schmidheini *et al.*, Continuous directed evolution of a compact CjCas9 variant with broad PAM compatibility. *Nat. Chem. Biol.* **20**, 333–343 (2024). doi: 10.1038/s41589-023-01427-x; pmid: 37735239
70. E. Zhang, M. E. Neugebauer, N. A. Krasnow, D. R. Liu, Phage-assisted evolution of highly active cytosine base editors with enhanced selectivity and minimal sequence context preference. *Nat. Commun.* **15**, 1697 (2024). doi: 10.1038/s41467-024-45969-7; pmid: 38402281
71. R. J. Sarnovsky, E. W. May, N. L. Craig, The Tn7 transposase is a heteromeric complex in which DNA breakage and joining activities are distributed between different gene products. *EMBO J.* **15**, 6348–6361 (1996). doi: 10.1002/j.1460-2075.1996.tb01024.x; pmid: 8947057
72. J. E. Peters, Tn7. *Microbiol. Spectr.* **2**, 2.5.27 (2014). doi: 10.1128/microbiolspec.MDNA3-0010-2014; pmid: 26104363
73. F. T. Hoffmann *et al.*, Selective TnsC recruitment enhances the fidelity of RNA-guided transposition. *Nature* **609**, 384–393 (2022). doi: 10.1038/s41586-022-05059-4; pmid: 36002573
74. J. Abramson *et al.*, Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024). doi: 10.1038/s41586-024-07487-w; pmid: 38718835
75. S. E. Halford, J. F. Marko, How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.* **32**, 3040–3052 (2004). doi: 10.1093/nar/gkh624; pmid: 15178741
76. M. W. G. Walker, S. E. Klompe, D. J. Zhang, S. H. Sternberg, Novel molecular requirements for CRISPR RNA-guided transposition. *Nucleic Acids Res.* **51**, 4519–4535 (2023). doi: 10.1093/nar/gkad270; pmid: 37078593
77. S. Wang, R. Siddique, M. C. Hall, P. A. Rice, L. Chang, Structure of TnsABCD transpososome reveals mechanisms of targeted DNA transposition. *Cell* **187**, 6865–6881.e16 (2024). doi: 10.1016/j.cell.2024.09.023; pmid: 39383864
78. G. D. Lampe, A. R. Liang, D. J. Zhang, I. S. Fernández, S. H. Sternberg, Structure-guided engineering of type I-F CASTs for targeted gene insertion in human cells. *bioRxiv*, 2024.2009.2019.613948 [Preprint] (2024); .doi: 10.1101/2024.09.19.613948
79. C. J. Tou, B. Orr, B. P. Kleinstiver, Precise cut-and-paste DNA insertion using engineered type V-K CRISPR-associated transposases. *Nat. Biotechnol.* **41**, 968–979 (2023). doi: 10.1038/s41587-022-01574-x; pmid: 36593413
80. E. W. May, N. L. Craig, Switching from cut-and-paste to replicative Tn7 transposition. *Science* **272**, 401–404 (1996). doi: 10.1126/science.272.5260.401; pmid: 8602527
81. G. Giannoukos *et al.*, UDiTaS™, a genome editing detection method for indels and genome rearrangements. *BMC Genomics* **19**, 212 (2018). doi: 10.1186/s12864-018-4561-9; pmid: 29562890
82. R. Žedaveinytė *et al.*, Antagonistic conflict between transposon-encoded introns and guide RNAs. *Science* **385**, eadm8189 (2024). doi: 10.1126/science.adm8189; pmid: 38991068
83. S. Q. Tsai *et al.*, GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–197 (2015). doi: 10.1038/nbt.3117; pmid: 25513782
84. J. T. George *et al.*, Mechanism of target site selection by type V-K CRISPR-associated transposases. *Science* **382**, eadj8543 (2023). doi: 10.1126/science.adj8543; pmid: 37972161
85. J. L. Doman, A. Raguram, G. A. Newby, D. R. Liu, Evaluation and minimization of Cas9-independent off-target DNA editing by cytosine base editors. *Nat. Biotechnol.* **38**, 620–628 (2020). doi: 10.1038/s41587-020-0414-6; pmid: 32042165
86. S. Banskota *et al.*, Engineered virus-like particles for efficient in vivo delivery of therapeutic proteins. *Cell* **185**, 250–265.e16 (2022). doi: 10.1016/j.cell.2021.12.021; pmid: 35021064
87. M. An *et al.*, Engineered virus-like particles for transient delivery of prime editor ribonucleoprotein complexes in vivo. *Nat. Biotechnol.* **42**, 1526–1537 (2024). doi: 10.1038/s41587-023-02078-y; pmid: 38191664
88. H. A. Rees *et al.*, Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. *Nat. Commun.* **8**, 15790 (2017). doi: 10.1038/ncomms15790; pmid: 28585549
89. W. H. Yeh, H. Chiang, H. A. Rees, A. S. B. Edge, D. R. Liu, In vivo base editing of post-mitotic sensory cells. *Nat. Commun.* **9**, 2184 (2018). doi: 10.1038/s41467-018-04580-3; pmid: 29872041
90. K. Kim *et al.*, Highly efficient RNA-guided base editing in mouse embryos. *Nat. Biotechnol.* **35**, 435–437 (2017). doi: 10.1038/nbt.3816; pmid: 28244995
91. J. A. Zuris *et al.*, Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nat. Biotechnol.* **33**, 73–80 (2015). doi: 10.1038/nbt.3081; pmid: 25357182
92. R. S. Isaac *et al.*, Nucleosome breathing and remodeling constrain CRISPR-Cas9 function. *eLife* **5**, e13450 (2016). doi: 10.7554/eLife.13450; pmid: 27130520
93. X. Li *et al.*, Chromatin context-dependent regulation and epigenetic manipulation of prime editing. *Cell* **187**, 2411–2427.e25 (2024). doi: 10.1016/j.cell.2024.03.020; pmid: 38608704
94. R. Sharma *et al.*, In vivo genome editing of the albumin locus as a platform for protein replacement therapy. *Blood* **126**, 1777–1784 (2015). doi: 10.1182/blood-2014-12-615492; pmid: 26297739

95. A. C. Nathwani *et al.*, Long-term safety and efficacy of factor IX gene therapy in hemophilia B. *N. Engl. J. Med.* **371**, 1994–2004 (2014). doi: 10.1056/NEJMoa1407309; pmid: 25409372
96. P. Simioni *et al.*, X-linked thrombophilia with a mutant factor IX (factor IX Padua). *N. Engl. J. Med.* **361**, 1671–1675 (2009). doi: 10.1056/NEJMoa0904377; pmid: 19846852
97. J. Eyquem *et al.*, Targeting a CAR to the TRAC locus with CRISPR/Cas9 enhances tumour rejection. *Nature* **543**, 113–117 (2017). doi: 10.1038/nature21405; pmid: 28225754
98. R. J. Brentjens *et al.*, Eradication of systemic B-cell tumors by genetically targeted human T lymphocytes co-stimulated by CD80 and interleukin-15. *Nat. Med.* **9**, 279–286 (2003). doi: 10.1038/nm827; pmid: 12579196
99. L. M. Lombardi *et al.*, An RNA interference screen identifies druggable regulators of MeCP2 stability. *Sci. Transl. Med.* **9**, eaaf7588 (2017). doi: 10.1126/scitranslmed.aaf7588; pmid: 28835516
100. M. J. Osborn *et al.*, Base Editor Correction of COL7A1 in Recessive Dystrophic Epidermolysis Bullosa Patient-Derived Fibroblasts and iPSCs. *J. Invest. Dermatol.* **140**, 338–347.e5 (2020). doi: 10.1016/j.jid.2019.07.701; pmid: 31437443
101. J. L. Doman, A. A. Sousa, P. B. Randolph, P. J. Chen, D. R. Liu, Designing and executing prime editing experiments in mammalian cells. *Nat. Protoc.* **17**, 2431–2468 (2022). doi: 10.1038/s41596-022-00724-4; pmid: 35941224
102. B. P. Dhungel, C. G. Bailey, J. E. J. Rasko, Journey to the Center of the Cell: Tracing the Path of AAV Transduction. *Trends Mol. Med.* **27**, 172–184 (2021). doi: 10.1016/j.molmed.2020.09.010; pmid: 33071047
103. J. U. Park *et al.*, Structures of the holo CRISPR RNA-guided transposon integration complex. *Nature* **613**, 775–782 (2023). doi: 10.1038/s41586-022-05573-5; pmid: 36442503
104. F. Tenjo-Castaño *et al.*, Structure of the TnsB transposase-DNA complex of type V-K CRISPR-associated transposon. *Nat. Commun.* **13**, 5792 (2022). doi: 10.1038/s41467-022-33504-5; pmid: 36184667
105. J. R. Fuller, P. A. Rice, Target DNA bending by the Mu transpososome promotes careful transposition and prevents its reversal. *eLife* **6**, e21777 (2017). doi: 10.7554/eLife.21777; pmid: 28177285
106. G. N. Barber, Innate immune DNA sensing pathways: STING, AIMII and the regulation of interferon production and inflammatory responses. *Curr. Opin. Immunol.* **23**, 10–20 (2011). doi: 10.1016/j.coi.2010.12.015; pmid: 21239155
107. J. R. Rybarski, K. Hu, A. M. Hill, C. O. Wilke, I. J. Finkelstein, Metagenomic discovery of CRISPR-associated transposons. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2112279118 (2021). doi: 10.1073/pnas.2112279118; pmid: 34845024
108. J. Liu *et al*., Integration of therapeutic cargo into the human genome with programmable type V-K CAST. *Nat Commun* **16**, 2427 (2025). doi: 10.1038/s41467-025-57416-2; pmid: 40082411
109. S. W. Smeal, M. A. Schmitt, R. R. Pereira, A. Prasad, J. D. Fisk, Simulation of the M13 life cycle I: Assembly of a genetically-structured deterministic chemical kinetic simulation. *Virology* **500**, 259–274 (2017). doi: 10.1016/j.virol.2016.08.017; pmid: 27644585
110. K. Clement *et al.*, CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019). doi: 10.1038/s41587-019-0032-3; pmid: 30809026
111. K. Clement, R. Farouni, D. E. Bauer, L. Pinello, AmpUMI: Design and analysis of unique molecular identifiers for deep amplicon sequencing. *Bioinformatics* **34**, i202–i210 (2018). doi: 10.1093/bioinformatics/bty264; pmid: 29949956
112. K. A. Datsenko, B. L. Wanner, One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6640–6645 (2000). doi: 10.1073/pnas.120163297; pmid: 10829079
113. S. A. Scholz *et al.*, High-Resolution Mapping of the Escherichia coli Chromosome Reveals Positions of High and Low Transcription. *Cell Syst.* **8**, 212–225.e9 (2019). doi: 10.1016/j.cels.2019.02.004; pmid: 30904377
114. H. Li, Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018). doi: 10.1093/bioinformatics/bty191; pmid: 29750242
115. N. Liscovitch-Brauer *et al.*, Profiling the genetic determinants of chromatin accessibility with scalable single-cell CRISPR screens. *Nat. Biotechnol.* **39**, 1270–1277 (2021). doi: 10.1038/s41587-021-00902-x; pmid: 33927415
116. I. P. Witte *et al.*, Programmable gene insertion in human cells with a laboratory-evolved CRISPR-associated transposase, Version v1, Zenodo (2025); doi: 10.5281/zenodo.14847695
117. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12 (2011). doi: 10.14806/ej.17.1.200
118. T. Smith, A. Heger, I. Sudbery, UMI-tools: Modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res.* **27**, 491–499 (2017). doi: 10.1101/gr.209601.116; pmid: 28100584
119. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012). doi: 10.1038/nmeth.1923; pmid: 22388286
120. T. Durfee *et al.*, The complete genome sequence of Escherichia coli DH10B: Insights into the biology of a laboratory workhorse. *J. Bacteriol.* **190**, 2597–2606 (2008). doi: 10.1128/JB.01695-07; pmid: 18245285
121. A. Karabacak Calviello, A. Hirsekorn, R. Wurmus, D. Yusuf, U. Ohler, Reproducible inference of transcription factor footprints in ATAC-seq and DNase-seq datasets using protocol-specific bias modeling. *Genome Biol.* **20**, 42 (2019). doi: 10.1186/s13059-019-1654-y; pmid: 30791920
122. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010). doi: 10.1093/bioinformatics/btq033; pmid: 20110278

ACKNOWLEDGMENTS

We thank A. Raguram, J. Doman, S. Pandey, D. Gao, and S. DeCarlo for helpful discussions. **Funding:** I.P.W. was supported by the US NSF Graduate Research Fellowship. R.T.K. was supported by NIH grant F31HL167530 from the National Heart, Lung, and Blood Institute. M.J.O, A.N.M., and J.T. are supported by NIH R01AR063070. S.H.S. was supported by NIH grants DP2HG011650 and R01EB027793, a Pew Biomedical Scholarship, a Sloan Research Fellowship, an Irma T. Hirschl Career Scientist Award, and a generous startup package from the Columbia University Irving Medical Center Dean's Office and the Vagelos Precision Medicine Fund. D.R.L. was supported by NIH grants R01EB031172, R01EB027793, RM1HG009490, R35GM118062, and HHMI. **Author contributions:** I.P.W., S.M.M., S.H.S., and D.R.L. conceived the project. S.M.M. designed the initial transposase evolution circuit, with input from D.R.G. and P.L.H.V. I.P.W. and S.M.M. performed evolution and characterization experiments to validate initial selection designs. I.P.W. performed transposase evolutions, optimized selection circuits, characterized and optimized evolved transposase variants, characterized evoCAST, and designed therapeutic proof-of-concept experiments. G.D.L. screened engineered Cas8 variants, screened QCascade NLS architectures, performed UDiTaS, characterized evoCAST, and assessed long-term persistence of edited cells. S.E. performed QCascade evolutions, engineered Cas7, characterized evoCAST, and performed therapeutic proof-of-concept experiments. K.N.B. performed HeLa and K562 cell experiments. A.N.M. and M.J.O. performed primary human fibroblast cell experiments. J.T. obtained RDEB fibroblasts and provided funding and support of A.N.M. and M.J.O. R.T.K. performed preliminary experiments assessing persistence of edited cells. O.G.S. performed experiments assessing transposase variants with combinations of evolved mutations. A.T.C. performed TnsC reversion analyses in HEK293T cells. D.R.L. and S.H.S. supervised the research and experimental design. I.P.W., G.D.L., S.H.S., and D.R.L. wrote the manuscript with input from all authors. **Competing interests:** The authors have filed patent applications related to this work. D.R.L. is a cofounder, consultant, and/or equity holder of Beam Therapeutics, Prime Medicine, Pairwise Plants, nChroma Bio, and Exo Therapeutics. S.H.S. is a cofounder and scientific adviser to Dahlia Biosciences, a scientific adviser to CrisprBits and Prime Medicine, and an equity holder in Dahlia Biosciences and CrisprBits. **Data and materials availability:** Plasmids used in this study are available through AddGene (depositor: D.R.L.). Sequencing files can be accessed through the National Center for Biotechnology Information Sequence Read Archive (accession no. PRJNA1224812). Custom scripts for analysis, PyMOL sessions for AlphaFold3-predicted structures, uncropped Western blot images, example gating for flow cytometry, and ddPCR data can be found on Zenodo (*116*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adt5199
Supplementary Text; Figs. S1 to S32; Tables S1 to S13; References (*123*–*131*); MDAR Reproducibility Checklist

Submitted 29 September 2024; accepted 23 February 2025

10.1126/science.adt5199